Solving Nonlinear Multicommodity Flow Problems by the Proximal Chebychev Center Cutting Plane Algorithm

Adam Ouorou∗


Abstract

The recent algorithm proposed in [15] (called pc3pa) for convex nonsmooth optimization, is specialized for applications in telecommunications on some nonlinear multicommodity flows problems. In this context, the objective function is additive and this property could be exploited for a better performance.

Keywords: nondifferentiable optimization, network flows, large-scale convex programming, routing problems in telecommunications.

1 Introduction

We are concerned with the convex problem

\[
\min_{x \in \mathbb{R}^n} f(x),
\]

where \( f \) is a separable convex function, i.e. of the form

\[
f(x) = \sum_{j \in J} f_j(x),
\]

each \( f_j : \mathbb{R}^n \to \mathbb{R} \) being a convex function, \( j \in J \triangleq \{1, \ldots, m\} \). A wide variety of convex programming can be modelled as (1), in particular the dual of a convex constrained separable problem (see [1, 6, 14]). For instance, consider the network utility maximization problem defined in a digraph \( G = (V, E, \gamma) \) with \( |E| = n, \gamma \in \mathbb{R}^n_+ \), supporting a set of \( \kappa \) demands represented by \( \kappa \) commodities between OD pairs \((O_p, T_p)\) with requirement \( d_p \geq 0 \) for \( p = 1, \ldots, \kappa \) (see [8])

\[
\begin{align*}
\min & \sum_{j=1}^n \varphi(y_j) - \sum_{p=1}^\kappa \psi_p(w_p) \\
\text{s.t.} & \sum_{p=1}^\kappa \sum_{s \in S(p)} \pi_p^s(j)x_p^s \leq y_j, \; j = 1, \ldots, n, \\
& \sum_{s \in S(p)} x_p^s = w_p, \; p = 1, \ldots, \kappa \\
& 0 \leq y_j \leq \gamma_j, \; j = 1, \ldots, n, \\
& x_p^s \geq 0, \; w_p \geq 0, \; p = 1, \ldots, \kappa,
\end{align*}
\]

where \( w_p \geq 0 \) be a real variable representing a rate allocated to the \( p^{th} \) OD pair, and \( S(p) \) be a given set of paths to support this amount. The rate \( w_p \) has utility \( \psi_p(w_p) \) for commodity \( p \). Utility \( \psi_p \) is assumed
to be an increasing strictly concave and continuously differentiable function. Let \( u \in \mathbb{R}^n \) and \( q_p \in \mathbb{R} \) be the dual multipliers associated to the capacity and demand constraints respectively. The dual problem is

\[
\min \sum_{j=1}^{n} \phi_j^*(u_j) + \sum_{p=1}^{\kappa} (-\psi_p)^*(-q_p)
\]

\[
s.t. \sum_{j=1}^{n} \pi_p^*(j)u_j \geq q_p, \ p = 1, \ldots, \kappa, \ s \in S(p),
\]

\[
u \geq 0, \ q_p \in \mathbb{R}, \ p = 1, \ldots, \kappa,
\]

which is of the form (1), ignoring the linear constraints. Our development can be extended to handle linear constraints with no harm.

Going back to the class of problem (1), if \( f \) is viewed as a single function, there is nothing new so far, the algorithm proposed in [15] can be applied as is. We could expect better performance by exploiting the additive structure of \( f \). This property is exploited in most of nonsmooth methods [4, 6, 7, 9] with in general no big deal. The specific form of \( pc^3pa \) algorithm needs a development.

2 The separable version of \( pc^3pa \)

It is well-known that any of the convex functions \( f_j \) can be written as the envelope of its supporting hyperplane, that is,

\[
f_j(x) = \sup \{f_j(z) + \langle g_j, x - z \rangle, \ z \in \mathbb{R}^n, \ g_j \in \partial f_j(z)\}.
\]

Hence,

\[
f(x) = \sum_{j \in J} \sup \{f_j(u) + \langle g_j, x - z \rangle, \ z \in \mathbb{R}^n, \ g_j \in \partial f_j(z)\}.
\]

Given some level point \( \bar{x} \in \mathbb{R}^n \), \( f(\bar{x}) \) is an upper bound to (1), and it is clear that the part of the epigraph that is below \( f(\bar{x}) \) defines a (level) set

\[
X_{\bar{x}} = \left\{ (x, r) \in \mathbb{R}^n \times \mathbb{R}^m : \langle e, r \rangle \leq f(\bar{x}), \ f_j(z) + \langle g_j, x - z \rangle \leq r_j, \ z \in \mathbb{R}^n, \ g_j \in \partial f_j(z), \ j \in J \right\}
\]

that contains the optimal set \( X^* \times \{f^*\} \) (\( e \in \mathbb{R}^m \) being the vector of all ones). Computing the Chebychev center of \( X_{\bar{x}} \) amounts to solving

\[
\begin{array}{l}
\min \nu \\
s.t. \quad \langle e, r \rangle - f(\bar{x}) \leq \nu, \\
\quad \frac{\langle g_j, x - z \rangle + f_j(z) - r_j}{\sqrt{1 + \|g_j\|^2}} \leq \nu, \ z \in \mathbb{R}^n, \ g_j \in \partial f_j(z), \ j \in J, \\
\quad \nu \in \mathbb{R}, \ r \in \mathbb{R}^m, \ x \in \mathbb{R}^n.
\end{array}
\]

where \( \nu \) represents the opposite of the radius of the sphere inscribed in \( X_{\bar{x}} \). We can not eliminate variable \( r \) as in the case \( m = 1 \) (which we refer to the aggregate case in the sequel), although we can show that the first constraint in (5) is tight at optimality as in [15]).
For any $x \in \mathbb{R}^n$, let $r^x$ be the vector of $\mathbb{R}^m$ whose $j^{th}$ component is $f_j(x)$, note that $\langle e, r^x \rangle = f(x)$.

Now, define the function $\psi_{\bar{x}}$ by

$$
\psi_{\bar{x}}(x, r) = \sup \left\{ \frac{\langle e, r - r^x \rangle}{\sqrt{m}}, \frac{\langle g_j, x - z \rangle + f_j(z) - r_j}{\sqrt{1 + \|g_j\|^2}} : (z, r) \in \mathbb{R}^n \times \mathbb{R}^m, g_j \in \partial f_j(z), j \in J \right\}.
$$

Then (5) is equivalent to

$$
\min_{(x, r) \in \mathbb{R}^n \times \mathbb{R}^m} \psi_{\bar{x}}(x, r) \quad (6)
$$

As in [15], we can devise an abstract algorithm for the solution of (1) as follows.

**Algorithm 2.1**

0. Choose some $\bar{x} \in \mathbb{R}^n$.
1. Solve (6) to get $(x_{\bar{x}}, r_{\bar{x}})$.
2. If $\psi_{\bar{x}}(x_{\bar{x}}, r_{\bar{x}}) = 0$ stop: $\bar{x}$ solves (1).
3. Set $\bar{x} = x_{\bar{x}}$ and loop to 1.

Solving (6) is equivalent to solving the problem

$$
\min_{(x, r) \in \mathbb{R}^n \times \mathbb{R}^m} \Phi_{\bar{x}}(x, r)
$$

where $\Phi_{\bar{x}}$ is the Moreau-Yosida regularization of $\psi_{\bar{x}}$ associated with a parameter $\mu > 0$, defined by

$$
\Phi_{\bar{x}}(x, r) = \min_{(z, v) \in \mathbb{R}^n} \left\{ \psi_{\bar{x}}(z, v) + \frac{\mu}{2} \| (z, v) - (x, r) \|^2 \right\}. \quad (7)
$$

Note that $\Phi_{\bar{x}}$ is continuously differentiable. The proximal point algorithm for minimizing $\Phi_{\bar{x}}$ is as follows.

**Algorithm 2.2**

0. Choose an initial guess $(y^1, w^1)$ (for instance $(\bar{x}, r_{\bar{x}})$) and set $j = 1$.
1. Compute

$$
(\tilde{y}^j, \tilde{w}^j) = \arg \min_{(z, v) \in \mathbb{R}^n} \left\{ \psi_{\bar{x}}(z, v) + \frac{\mu}{2} \| (z, v) - (y^j, w^j) \|^2 \right\}
$$

2. If $(y^j, w^j) = (\tilde{y}^j, \tilde{w}^j)$, stop: $(x_{\bar{x}}, r_{\bar{x}}) = (y^j, w^j)$. Otherwise set $(y^{j+1}, w^{j+1}) = (\tilde{y}^j, \tilde{w}^j)$.
3. Increase $j$ by 1 and loop to Step 1.

The solution in Step 1 is the proximal point of $(y^j, w^j)$ w.r.t. $\psi_{\bar{x}}$. Because the level point $\bar{x}$ drives Algorithm 2.1 (only $\bar{x}$ and the optimal value in (6) are important in this algorithm), in place of the above proximal point algorithm, we shall use a partial proximal point algorithm [2] by replacing the minimization problem is Step 1 by

$$
(\tilde{y}^j, \tilde{w}^j) = \arg \min_{(z, v) \in \mathbb{R}^n} \left\{ \psi_{\bar{x}}(z, v) + \frac{\mu}{2} \| z - y^j \|^2 \right\}. \quad (8)
$$

There is no need to involve variable $r$ in the quadratic term which would have preclude the minimization of variable $\nu$. The quadratic suproblem (8) may be solved as proposed in [15] using a bundle strategy.
In this aim, consider the following lower approximation to \( \psi \) obtained with a set of sample points \( z^i, g_j^i \in \partial f(z^i), \ i \in I^k \)

\[
\tilde{\psi}_{x,k}(x, r) = \max \left\{ \frac{\langle e, r \rangle - f(\bar{x})}{\sqrt{m}}, \frac{(g_j^i, x - z^i) + f_j(z^i) - r_j}{\sqrt{1 + \|g_j^i\|^2}}, \ j \in J, \ i \in I^k \right\}.
\]

(9)

Then, \( \tilde{\psi}_{x,k}(x, r) \leq \psi_{x,k}(x, r) \) for any \((x, r) \in \mathbb{R}^n \times \mathbb{R}^m\) and a bundle scheme could be used to solve (8). Using the same arguments and development of [15], an approximate stopping criterion in Step 2 could be replaced by

\[
f(z^{k+1}) \leq f(\bar{x}) + \kappa (m + \sqrt{m}) \tilde{\psi}_{x,k}(z^{k+1}, v^{k+1}),
\]

(10)

for some \(0 < \kappa < 1\).

For a fixed proximal center \( \bar{x} \), the QP subproblem in the bundle scheme is as follows

\[
\min \quad \nu + \frac{\mu}{2} \|x - \bar{x}\|^2 \\
\text{s.t.} \quad \frac{1}{\sqrt{m}} ((\mathbf{e}, r) - f(\bar{x})) \leq \nu, \\
(\tilde{g}_j^i, x - \bar{x}) - \alpha_{i,k} + \gamma_{ij}(f_j(\bar{x}) - r_j) \leq \nu, \ j \in J, \ i \in I^k,
\]

(11)

where

\[
\begin{align*}
\alpha(z, z^i) &= f_j(z^i) - \left[f_j(z^i) + (g_j^i, x - z^i) \right] \text{ is the linearization error between } z^i \text{ and } \bar{x}, \text{ (it avoids storing } z^i), \\
\gamma_{ij} &= (1 + \|g_j^i\|^2)^{-\frac{1}{2}}, \ i \in I^k, \ j = 1, \ldots, m, \\
\tilde{g}_j^i &= \gamma_{ij}g_j^i \text{ and } \alpha_{i,k} = \gamma_{ij} \alpha(z, z^i) \text{ are respectively the scaled subgradients and scaled linearization errors.}
\end{align*}
\]

Let \( \lambda_0 \geq 0 \) and \( \lambda_{ij} \geq 0, \ i \in I^k, \ j \in J \) be the Lagrange multipliers of (11). The Lagrangian function is

\[
L(r, x, \lambda) = (1 - \lambda_0 - \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda_{ij})\nu + \frac{\mu}{2} \|x - \bar{x}\|^2 + \frac{\lambda_0}{\sqrt{m}} ((\mathbf{e}, r) - f(\bar{x})) + \\
\sum_{j=1}^{m} \sum_{i=1}^{k} \lambda_{ij} \left[ (\tilde{g}_j^i, x - \bar{x}) - \alpha_{i,k} + \gamma_{ij}(r_j^x - r_j) \right]
\]

which minimization w.r.t. \( \nu, r \) and \( x \) yields

\[
1 - \lambda_0 - \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda_{ij} = 0, \tag{12}
\]

\[
\frac{\lambda_0}{\sqrt{m}} = \sum_{i=1}^{k} \lambda_{ij} \gamma_{ij}, \ j = 1, \ldots, m, \tag{13}
\]

\[
x = \bar{x} - \frac{1}{\mu} \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda_{ij} \tilde{g}_j^i. \tag{14}
\]
Substituting this into the Lagrangian gives the dual problem

\[
\min \frac{1}{2\mu} \left\| \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda_{ij} \tilde{g}_j \right\|^2 + \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda_{ij} \tilde{a}_{ij}^k \\
\text{s.t.} \quad \sqrt{m} \sum_{i=1}^{k} \lambda_{ij} \gamma_{ij} + \sum_{i=1}^{k} \sum_{j'=1}^{m} \lambda_{ij'} = 1, \quad j \in J,
\]

(15)

Let \( \lambda^{k}_{ij}, \ i \in I_k, \ j = 1, \ldots, m, \) be its optimal solution of (15), and define

\[
g_{a,j}^k = \sum_{i=1}^{k} \lambda^{k}_{ij} \tilde{g}_j, \quad \alpha_{a,j}^k = \sum_{i=1}^{k} \lambda^{k}_{ij} \alpha_{i,k}, \quad g_a^k = \sum_{j=1}^{m} g_{a,j}^k \quad \text{and} \quad \alpha_a^k = \sum_{j=1}^{m} \alpha_{a,j}^k.
\]

(16)

The optimal solution of (11) is as follows

\[
\nu^k = -\frac{1}{\mu} \left\| g_a^k \right\|^2 - \alpha_a^k, \quad z^{k+1} = \bar{x} - \frac{1}{\mu} g_a^k.
\]

(17)

Using the subgradient inequalities for each component function, we have for each \( j \) and all \( x \in \mathbb{R}^n \)

\[
\lambda^{k}_{ij} \gamma_{ij} f_j(x) \geq \lambda^{k}_{ij} \gamma_{ij} (x^k) + (\lambda^{k}_{ij} \tilde{g}_j, x - x^k) - \lambda^{k}_{ij} \alpha_{i,k}, \quad i \in I^k.
\]

Summing over \( j \in J \) and \( i \in I^k \), we get

\[
\sum_{j=1}^{m} \sum_{i=1}^{k} \lambda^{k}_{ij} \gamma_{ij} f_j(x) \geq \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda^{k}_{ij} \gamma_{ij} f_j(\bar{x}) + \left( \sum_{j=1}^{m} g_{a,j}^k, x - \bar{x} \right) - \sum_{j=1}^{m} \alpha_a^k.
\]

Let \( \lambda^{k}_0 = 1 - \sum_{j=1}^{m} \sum_{i=1}^{k} \lambda^{k}_{ij} > 0 = \sqrt{m} \sum_{i=1}^{k} \lambda^{k}_{ij} \gamma_{ij} \) for any \( j \). We have \( \lambda^{k}_0 > 0 \) since \( \lambda^{k}_0 = 0 \) implies that all the \( \lambda^{k}_{ij} \) are null (because \( \gamma_{ij} > 0 \)) which contradicts (12). Let \( \gamma_a^k \) be the inverse of the common value of the \( \sum_{i=1}^{k} \lambda^{k}_{ij} \gamma_{ij} \) for \( j \in J \) i.e.

\[
\gamma_a^k = \frac{\sqrt{m}}{\lambda^{k}_0}.
\]

(18)

We get from (13) and (17) that,

\[
f(x) \geq f(\bar{x}) + (\gamma_a^k g_a^k, x - \bar{x}) - \gamma_a^k \alpha_a^k,
\]

(19)

or equivalently \( \gamma_a^k g_a^k \in \partial \kappa \alpha_a^k f(\bar{x}) \). But \( \nu^k = 0 \) implies \( g_a^k = 0 \) and \( \alpha_a^k = 0 \) from (17).

Putting all things together, we now state the separable version of the proximal algorithm proposed in [15] as follows.

**Algorithm** Separable Proximal Chebychev Center Cutting Plane Algorithm (spc³pa)

0. Select the stopping tolerance \( \varepsilon \), the parameter \( 0 < \kappa < 1 \), an initial point \( z^1 \in \mathbb{R}^n \). Compute \( f(z^1), \ g_j^1 \in \partial f(z^1), \ j \in J \). Set \( x^1 = z^1 \) and \( k = 1 \).

1. If \( \sum_{j=1}^{m} g_j^k = 0 \), terminate.
2. Solve (15) to obtain $\lambda_i^k, \ i \in I^k$. Compute $g_a^k, \alpha_a^k$ and $\gamma_a^k$ (cf (16) and (18)). Set

$$z^{k+1} = x^k - \frac{1}{\mu_k}g_a^k, \ \sigma^k = \frac{1}{\mu_k}g_a^k + \alpha_a^k.$$ (20)

3. If $\gamma_a^k\sigma^k \leq \varepsilon$, terminate.

4. Compute $f_j(z^{k+1}), \ g_{ja}^{k+1} \in \partial f_j(z^{k+1}), \ j \in J$.

5. If $f(z^{k+1}) \leq f(x^k) - \kappa(m + \sqrt{m})\sigma^k$, set $x^{k+1} = z^{k+1}$. Otherwise $x^{k+1} = x^k$.

6. Increase $k$ by 1 and loop to Step 1.

The stopping criterion is based on the fact that

$$f(x^k) \leq f(x) + \gamma_a^k\|g_a^k\||x - x^k| + \gamma_a^k\alpha_a^k, \ \text{for all} \ x \in \mathbb{R}^n.$$ which results from (19) using Cauchy-Schwartz inequality. The quadratic subproblem could be solved using CPLEX even if specialized algorithms exist [10, 5]. In case the problem involves linear constraints (as in (3)), these constraints are simply added to the quadratic subproblem.

References


