

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

Adam Ouorou*

*Orange Labs R&D/CORE-MCN,

38-40 rue du Général Leclerc, 92794 Issy-Les-Moulineaux cedex 9, France.

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

$$\min_{x \in \mathbb{R}^n} f(x), \quad (1)$$

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

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

each $f_j : \mathbb{R}^n \rightarrow \mathbb{R}$ being a convex function, $j \in J \triangleq \{1, \dots, m\}$. 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 κ demands represented by κ commodities between OD pairs (O_p, T_p) with requirement $d_p \geq 0$ for $p = 1, \dots, \kappa$ (see [8])

$$\begin{aligned} \min \quad & \sum_{j=1}^n \varphi(y_j) - \sum_{p=1}^{\kappa} \psi_p(w_p) \\ \text{s.t.} \quad & \sum_{p=1}^{\kappa} \sum_{s \in S(p)} \pi_p^s(j) x_p^s \leq y_j, \quad j = 1, \dots, n, \\ & \sum_{s \in S(p)} x_p^s = w_p, \quad p = 1, \dots, \kappa \\ & 0 \leq y_j \leq \gamma_j, \quad j = 1, \dots, n, \\ & x_p^s \geq 0, \quad w_p \geq 0, \quad p = 1, \dots, \kappa, \end{aligned} \quad (2)$$

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 ψ_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

$$\begin{aligned}
\min \quad & \sum_{j=1}^n \varphi_j^*(u_j) + \sum_{p=1}^{\kappa} (-\psi_p)^*(-q_p) \\
\text{s.t.} \quad & \sum_{j=1}^n \pi_p^s(j) u_j \geq q_p, \quad p = 1, \dots, \kappa, \quad s \in S(p), \\
& u \geq 0, \quad q_p \in \mathbb{R}, \quad p = 1, \dots, \kappa,
\end{aligned} \tag{3}$$

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 pc³pa algorithm needs a development.

2 The separable version of pc³pa

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

$$\begin{aligned}
X_{\bar{x}} = \left\{ (x, r) \in \mathbb{R}^n \times \mathbb{R}^m : \langle \mathbf{e}, r \rangle \leq f(\bar{x}), \quad f_j(z) + \langle g_j, x - z \rangle \leq r_j, \quad z \in \mathbb{R}^n, \right. \\
\left. g_j \in \partial f_j(z), \quad j \in J \right\}
\end{aligned} \tag{4}$$

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

$$\begin{aligned}
\min \quad & \nu \\
\text{s.t.} \quad & \frac{\langle \mathbf{e}, r \rangle - f(\bar{x})}{\sqrt{m}} \leq \nu, \\
& \frac{\langle g_j, x - z \rangle + f_j(z) - r_j}{\sqrt{1 + \|g_j\|^2}} \leq \nu, \quad z \in \mathbb{R}^n, \quad g_j \in \partial f_j(z), \quad j \in J,
\end{aligned} \tag{5}$$

$$\nu \in \mathbb{R}, \quad r \in \mathbb{R}^m, \quad x \in \mathbb{R}^n.$$

where ν 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 $\langle \mathbf{e}, r^x \rangle = f(x)$. Now, define the function $\psi_{\bar{x}}$ by

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

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 in 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 ν . The quadratic subproblem (8) may be solved as proposed in [15] using a bundle strategy.

In this aim, consider the following lower approximation to $\psi_{\bar{x}}$ obtained with a set of sample points $z^i, g_j^i \in \partial f(z^i)$, $i \in I^k$

$$\check{\psi}_{\bar{x},k}(x, r) = \max \left\{ \frac{\langle \mathbf{e}, r \rangle - f(\bar{x})}{\sqrt{m}}, \frac{\langle g_j^i, x - z^i \rangle + f_j(z^i) - r_j}{\sqrt{1 + \|g_j^i\|^2}}, j \in J, i \in I^k \right\}. \quad (9)$$

Then, $\check{\psi}_{\bar{x},k}(x, r) \leq \psi_{\bar{x},k}(x, r)$ 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

$$f(z^{k+1}) \leq f(\bar{x}) + \kappa(m + \sqrt{m})\check{\psi}_{\bar{x},k}(z^{k+1}, v^{k+1}), \quad (10)$$

for some $0 < \kappa < 1$.

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

$$\begin{aligned} \min \quad & \nu + \frac{\mu}{2} \|x - \bar{x}\|^2 \\ \text{s.t.} \quad & \frac{1}{\sqrt{m}} (\langle \mathbf{e}, r \rangle - f(\bar{x})) \leq \nu, \\ & \langle \tilde{g}_j^i, x - \bar{x} \rangle - \alpha_{i,k}^j + \gamma_{ij}(f_j(\bar{x}) - r_j) \leq \nu, \quad j \in J, i \in I^k, \\ & \nu \in \mathbb{R}, \quad r \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \end{aligned} \quad (11)$$

where

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

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

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

which minimization w.r.t. ν , r and x yields

$$1 - \lambda_0 - \sum_{j=1}^m \sum_{i=1}^k \lambda_{ij} = 0, \quad (12)$$

$$\frac{\lambda_0}{\sqrt{m}} = \sum_{i=1}^k \lambda_{ij} \gamma_{ij}, \quad j = 1, \dots, m, \quad (13)$$

$$x = \bar{x} - \frac{1}{\mu} \sum_{j=1}^m \sum_{i=1}^k \lambda_{ij} \tilde{g}_j^i. \quad (14)$$

Substituting this into the Lagrangian gives the dual problem

$$\begin{aligned}
\min \quad & \frac{1}{2\mu} \left\| \sum_{j=1}^m \sum_{i=1}^k \lambda_{ij} \tilde{g}_j^i \right\|^2 + \sum_{j=1}^m \sum_{i=1}^k \lambda_{ij} \tilde{\alpha}_{ij}^k \\
\text{s.t.} \quad & \sqrt{m} \sum_{i=1}^k \lambda_{ij} \gamma_{ij} + \sum_{i=1}^k \sum_{j'=1}^m \lambda_{ij'} = 1, \quad j \in J, \\
& \lambda_{ij} \geq 0, \quad i \in I^k, \quad j \in J.
\end{aligned} \tag{15}$$

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

$$g_{a,j}^k = \sum_{i=1}^k \lambda_{ij}^k \tilde{g}_j^i, \quad \alpha_{a,j}^k = \sum_{i=1}^k \lambda_{ij}^k \alpha_{i,k}^j, \quad g_a^k = \sum_{j=1}^m g_{a,j}^k \quad \text{and} \quad \alpha_a^k = \sum_{j=1}^m \alpha_{a,j}^k. \tag{16}$$

The optimal solution of (11) is as follows

$$\nu^k = -\frac{1}{\mu} \|g_a^k\|^2 - \alpha_a^k, \quad z^{k+1} = \bar{x} - \frac{1}{\mu} g_a^k. \tag{17}$$

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

$$\lambda_{ij}^k \gamma_{ij} f_j(x) \geq \lambda_{ij}^k \gamma_{ij} f_j(x^k) + \langle \lambda_{ij}^k \tilde{g}_j^i, x - x^k \rangle - \lambda_{ij}^k \alpha_{i,k}^j, \quad i \in I^k.$$

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

$$\sum_{j=1}^m \sum_{i=1}^k \lambda_{ij}^k \gamma_{ij} f_j(x) \geq \sum_{j=1}^m \sum_{i=1}^k \lambda_{ij}^k \gamma_{ij} f_j(\bar{x}) + \langle \sum_{j=1}^m g_{a,j}^k, x - \bar{x} \rangle - \sum_{j=1}^m \alpha_{a,j}^k.$$

Let $\lambda_0^k = 1 - \sum_{j=1}^m \sum_{i=1}^k \lambda_{ij}^k > 0 = \sqrt{m} \sum_{i=1}^k \lambda_{ij}^k \gamma_{ij}$ for any j . We have $\lambda_0^k > 0$ since $\lambda_0^k = 0$ implies that all the λ_{ij}^k are null (because $\gamma_{ij} > 0$) which contradicts (12). Let γ_a^k be the inverse of the common value of the $\sum_{i=1}^k \lambda_{ij}^k \gamma_{ij}$ for $j \in J$ i.e.

$$\gamma_a^k = \frac{\sqrt{m}}{\lambda_0^k}. \tag{18}$$

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

$$f(x) \geq f(\bar{x}) + \langle \gamma_a^k g_a^k, x - \bar{x} \rangle - \gamma_a^k \alpha_a^k, \tag{19}$$

or equivalently $\gamma_a^k g_a^k \in \partial_{\gamma_a^k \alpha_a^k} f(\bar{x})$. 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 ε , 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 λ_i^k , $i \in I^k$. Compute g_a^k , α_a^k and γ_a^k (cf (16) and (18)). Set

$$z^{k+1} = x^k - \frac{1}{\mu^k} g_a^k, \quad \sigma^k = \frac{1}{\mu^k} \|g_a^k\|^2 + \alpha_a^k. \quad (20)$$

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

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

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

$$f(x^k) \leq f(x) + \gamma_a^k \|g_a^k\| \|x - x^k\| + \gamma_a^k \alpha_a^k, \quad \text{for all } x \in \mathbb{R}^n.$$

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

- [1] W. Ben Ameur, A. Ouorou, *Mathematical models of the delay constrained routing problem*, Algorithmic Operations Research, **1**(2) (2006) pp. 94–103.
- [2] D. P. Bertsekas, P. Tseng, *Partial proximal minimization algorithms for convex programming*, SIAM Journal on Optimization, **4**(3) (1994) pp. 551–572.
- [3] S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.
- [4] J. Elzinga, T.G. Moore, *A central cutting plane algorithm for the convex programming problem*, Mathematical Programming **8** (1975) pp. 134–145.
- [5] A. Frangioni, *Solving semidefinite quadratic problems within nonsmooth optimization algorithms*, Computers & OR, **23**(11) (1996) pp. 1099–1118.
- [6] A. Frangioni, *Dual-ascent methods and multicommodity flow problems* Ph.D. Thesis, Università di Pisa, Italy, 1997.
- [7] J.-B. Hiriart-Urruty, C. Lemaréchal, *Convex Analysis and Minimization Algorithms*, Springer-Verlag, Berlin, 1993.
- [8] F. Kelly, *Charging and rate control for elastic traffic*, European Transactions on Telecommunications, **8** (1997) pp. 33–37.
- [9] K.C. Kiwiel, *Proximity control in bundle methods for convex nondifferentiable minimization*, Mathematical Programming, **46**(1) (1990) pp. 105–122.
- [10] K.C. Kiwiel, *A Cholesky dual method for proximal piecewise linear programming*, Numerische Mathematik **68**(3) (1994) pp. 325–340.
- [11] C. Lemaréchal, A. Ouorou, G. Petrou, *An efficient bundle algorithm for routing in telecommunication data networks*, Computational Optimization and Applications, (2008) to appear.
- [12] G.L. Nemhauser, W.B. Widhelm, *A modified linear program for columnar methods in mathematical programming*, Operations Research **19** (1971) pp. 1051–1060.

- [13] Y. Nesterov, *Introductory Lectures on Convex Programming: Basic Course*, Kluwer Academic Publishers, 2003.
- [14] A. Ouorou, P. Mahey, and J.-Ph. Vial, *A survey of algorithms for convex multicommodity flow problems*, *Management Science*, **46**(1) (2000) pp. 126–147.
- [15] A. Ouorou, *A Proximal cutting plane method using Chebychev center for nonsmooth convex optimization*, *Mathematical Programming* (2008) To appear.