DP Programming: optimality conditions and numerical methods

Adil Bagirov

Federation University Australia, Ballarat, Victoria, Australia

WoMBAT2018, Deakin University, November 29, 2018

Joint work with Julien Ugon

Outline

- Introduction
- Convex piecewise linear optimization
- DP Programming: local optimality conditions
- DP Programming: global optimality conditions
- DP Programming: numerical algorithms
- Constrained DP Programming
- Conclusions

Difference of convex (DC) programming

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a DC if it can be represented as a difference of two convex functions:

$$f(x) = f_1(x) - f_2(x)$$

where $f_1, f_2 : \mathbb{R}^n \to \mathbb{R}$ are convex functions.

DC programming problem:

$$\text{minimize} \quad f(x), \ x \in \mathbb{R}^n$$

subject to

$$h_i(x) = 0, i \in I, \quad g_j(x) \leq 0, \ j \in J.$$

Functions $f, h_i, i \in I, \ g_j, \ j \in J$ are DC functions.

Horst, Thoai, Tuy, An & Tao.

Branch & Bound, DCA (DC algorithms).

Difference of polyhedral (DP) programming

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a DP if it can be represented as a difference of two convex polyhedral functions:

$$f(x) = f_1(x) - f_2(x)$$

where functions f_1 and f_2 are convex polyhedral:

$$f_i(x) = \max_{j \in J_i} \varphi_{ij}(x)$$

DP programming problem:

$$\text{minimize} \quad f(x), \; x \in \mathbb{R}^n$$

subject to

$$h_i(x) = 0, i \in I, \quad g_j(x) \le 0, \; j \in J.$$

Functions $f, h_i, i \in I, \; g_j, \; j \in J$ are DP functions.

Introduction

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called a piecewise linear if there are finite number of sets $D_i \subset \mathbb{R}^n,\ i = 1, \ldots, m$ such that $f(x) = f_i(x),\ x \in D_i$ and the function f_i is affine.

A continuous piecewise linear function $f : \mathbb{R}^n \to \mathbb{R}$ can be represented as a max-min of affine functions:

$$f(x) = \max_{i \in I} \min_{j \in J_i} \varphi_{ij}(x).$$

$$\varphi_{ij}(x) = \langle a^{ij}, x \rangle + b_{ij}, \quad a^{ij} \in \mathbb{R}^n,\ b_{ij} \in \mathbb{R}.$$

Functions represented as a max-min of linear functions are DP:

$$f(x) = f_1(x) - f_2(x)$$

where

$$f_1(x) = \max_{i \in I} \left[\sum_{j \in J_i} \varphi_{ij}(x) + \sum_{k \in I, k \neq i} \max_{j \in J_k} \sum_{t \in J_k, t \neq j} \varphi_{ij}(x) \right],$$

$$f_2(x) = \sum_{i \in I} \max_{j \in J_i} \sum_{t \in J_i, t \neq j} \varphi_{ij}(x).$$

Introduction

There are many applications of such DP functions:

- Cluster analysis;
- Supervised data classification;
- Regression analysis;
- Clusterwise linear regression.

Introduction

Global Optimization: the cutting angle method (Rubinov, 1997; Bagirov and Rubinov, 2000). Let

$$l \in \mathbb{R}^n_+, l \neq 0,$$

$$I(l) = \{i = 1, \ldots, n : l_i > 0\},$$

$$S = \left\{ x \in \mathbb{R}^n_+ : \sum_{i=1}^{n} x_i = 1 \right\}.$$

$$h_j(x) = \max_{k \leq j} \min_{i \in I(l^k)} l_i^k x_i.$$

Main step in the cutting angle method is as follows:

$$\text{minimize} \quad h_j(x) \quad \text{subject to} \quad x \in S.$$

Introduction

We consider the case when the function f is represented as a difference of two maximum of linear functions:

$$f(x) = f_1(x) - f_2(x)$$

where

$$f_1(x) = \max_{i \in I_1} \varphi_{1i}(x), \quad f_2(x) = \max_{i \in I_2} \varphi_{2i}(x).$$

First we consider the unconstrained piecewise linear optimization problem:

$$\text{minimize} \quad f(x) \text{ subject to } \quad x \in \mathbb{R}^n$$

Convex piecewise linear optimization

Consider the function

$$f(x) = \max_{i \in I} \langle c^i, x \rangle, \quad c^i \in \mathbb{R}^n, \quad I = \{1, \ldots, m\}. \tag{1}$$

Its subdifferential is:

$$\partial f(x) = \text{co} \left\{ c^i, \ i \in R(x) \right\}, \ R(x) = \left\{ i \in I : \ \langle c^i, x \rangle = f(x) \right\}.$$

The necessary and sufficient optimality condition:

$$0_n \in \partial f(x).$$

Convex piecewise linear optimization

Demyanov (1968), Fisher, Northup and Shapiro (1975), Wolfe (1975), Grinold (1972), Dantzig and Wolfe (1961), Brooks and Geoffrion (1966), Eaves (1974), Hel, Wolfe and Growder (1974), Bazaraa, Goode and Rardin (1978), Kiwiel (1985).

Convex piecewise linear optimization

The minimization of the function (1) can be replaced by the following LP problem:

$$\text{minimize} \quad u$$

subject to

$$u \in \mathbb{R}, x \in \mathbb{R}^n,$$

$$\langle c^i, x \rangle \leq u, \;\; i \in I$$

Convex piecewise linear optimization

Let $\varepsilon > 0$ be given. $f^* = \inf f(x)$.

Algorithm 1 Minimization of convex piecewise linear functions.

Step 1. Select a starting point $x^1 \in \mathbb{R}^n$ and set $k := 1$.

Step 2. Compute $\bar{v}^k \in \mathbb{R}^n$ such that

$$\|\bar{v}^k\|^2 = \min\left\{\|z\|^2 : z \in \partial f(x^k)\right\}.$$

Step 3. If $\bar{v}^k = 0$ then the algorithm terminates.

Step 4. Set $x^{k+1} := x^k - \alpha_k \bar{v}^k$ where α_k is computed as follows:

$$\hat{R}(x^k, -\bar{v}^k) = \left\{ i \in R(x^k) : \ \langle c^i, -\bar{v}^k \rangle = \max_{v \in \partial f(x^k)} \langle v, -\bar{v}^k \rangle \right\},$$

$$\bar{R}(x^k, -\bar{v}^k) = \left\{ i \in I \setminus R(x^k) : \ \langle c^i, -\bar{v}^k \rangle > 0 \right\},$$

$$\alpha_k = \min \left\{ \frac{\langle c^i - c^j, x^k \rangle}{\langle c^i - c^j, \bar{v}^k \rangle}, \ i \in \hat{R}(x^k, -\bar{v}^k), \ j \in \bar{R}(x^k, -\bar{v}^k) \right\}.$$

Set $k = k + 1$ and go to Step 6.

Algorithm 1 is finite convergent.

Now consider more general DP functions:

$$f(x) = \max_{i \in I_1} \varphi_{1i}(x) - \max_{j \in I_2} \varphi_{2j}(x). \tag{2}$$

and DP programming problem:

$$\text{minimize} \quad f(x) \quad \text{subject to} \quad x \in \mathbb{R}^n. \tag{3}$$

Here

$$f_1(x) = \max_{i \in I_1} \varphi_{1i}(x), \quad f_2(x) = \max_{j \in I_2} \varphi_{2j}(x).$$

DP programming

M. Gaudioso, G. Giallombardo, G. Miglionico: Minimizing Piecewise-Concave Functions Over Polyhedra, Mathematics of Operations Research, 43(2), 2017.

L. Polyakova: On global unconstrained minimization of the difference of polyhedral functions, Journal of Global Optimization, 2011, Volume 50, Issue 2, pp 179–195.

Nguyen Thi Van HangNguyen Dong Yen, On the Problem of Minimizing a Difference of Polyhedral Convex Functions Under Linear Constraints, JOTA, 2016, Volume 171, Issue 2, pp 617642

DP programming

For the Clarke subdifferential $\partial f(x)$ we have

$$\partial f(x) \subset \partial f_1(x) - \partial f_2(x).$$

Different stationary points can be defined for Problem (3):

A point x^* is called an inf-stationary for the problem (3) if

$$\partial f_2(x^*) \subset \partial f_1(x^*). \tag{4}$$

A point x^* is called a Clarke stationary for the problem (3) if

$$0 \in \partial f(x^*). \tag{5}$$

Finally, a point x^* is called a critical point of the problem (3) if

$$\partial f_1(x^*) \cap \partial f_2(x^*) \neq \emptyset. \tag{6}$$

The function f can be rewritten as:

$$f(x) = \min_{j \in I_2} \max_{i \in I_1} \left\{ \varphi_{1i}(x) - \varphi_{2j}(x) \right\}.$$

Then the problem (3) can be replaced by $|I_2|$ convex problems of the form:

$$\text{minimize} \quad \max_{i \in I_1} \left\{ \varphi_{1i}(x) - \varphi_{2j}(x) \right\} \quad \text{subject to } x \in \mathbb{R}^n, \ j \in I_2.$$

DP programming

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called quasidifferentiable at a point x if it is locally Lipschitz continuous, directionally differentiable at this point and there exist convex, compact sets $\underline{\partial} f(x)$ and $\overline{\partial} f(x)$ such that:

$$f'(x,d) = \max_{u \in \underline{\partial} f(x)} \langle u, d \rangle + \min_{v \in \overline{\partial} f(x)} \langle v, d \rangle.$$

The pair of sets $D(x) = [\underline{\partial} f(x), \overline{\partial} f(x)]$ is called a quasidifferential of the function f at a point x (Demyanov, Rubinov, 1979).

For the DP function:

$$D(x) = [\partial f_1(x), -\partial f_2(x)].$$

DP Programming

Let $X \subset \mathbb{R}^n$ be an open set. A function f is defined on X and it is finite. We call this function codifferentiable at a point $x \in X$, if there exist convex compact sets $\underline{d}f(x) \subset \mathbb{R}^{n+1}$ and $\overline{d}f(x) \subset \mathbb{R}^{n+1}$ such that

$$f(x+\Delta) = f(x) + \Phi_x(\Delta) + o_x(\Delta)$$

where

$$\Phi_x(\Delta) = \max_{(a,v)\in \underline{d}f(x)} \left[a + \langle v, \Delta\rangle\right] + \min_{(b,w)\in \overline{d}f(x)} \left[b + \langle w, \Delta\rangle\right],$$

$$\frac{o_x(\alpha\Delta)}{\alpha} \longrightarrow_{\alpha\downarrow 0} 0, \quad \forall \Delta \in \mathbb{R}^n.$$

Here $a, b \in \mathbb{R}^1, \quad v, w \in \mathbb{R}^n$. We assume that co $\{x, x+\Delta\} \subset X$.

The pair $Df(x) = [\underline{d}f(x), \overline{d}f(x)]$ is called a codifferential, the set $\underline{d}f(x)$ - hypodifferential and the set $\overline{d}f(x)$ - hyperdifferential of the function f at x.

A function f is said to be continuously codifferentiable at a point $x \in X$ if it is codifferentiable in some neighborhood of this point and mappings $x \mapsto \underline{d}f(x)$, $x \mapsto \overline{d}f(x)$ are Hausdorf continuous.

The class of quasidifferentiable and codifferentiable functions coincide. (Demyanov, 1988).

DP programming

Consider the following maximum function

$$f(x) = \max_{i \in I} f_i(x),$$

where functions f_i are continuously differentiable. This function is hypodifferentiable and its hypodifferential is as follows

$$\underline{d}f(x) = \text{co}\,\{(a, v) \in \mathbb{R}^{n+1} : \ a = f_i(x) - f(x), \ v = \nabla f_i(x), \ i \in I\}.$$

Necessary condition for a minimum:

$$0_{n+1} \in [\underline{d}f(x) + [0, w]] \quad \forall \ [0, w] \in \overline{d}f(x).$$

Proposition 1 *Let f be a DP function. Then for any $x, y \in \mathbb{R}^n$*

$$f(y) - f(x) = \max_{(\eta,v)\in \underline{d}f(x)} [\eta + \langle v, y - x\rangle] + \min_{(\theta,w)\in \overline{d}f(x)} [\theta + \langle w, y - x\rangle]. \quad (7)$$

Corollary 1 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a DP function. Then at a point x there exists $\varepsilon > 0$ such that*

$$f(y) = f(x) + \max_{v\in \underline{\partial} f(x)} \langle v, y - x\rangle + \min_{w\in \overline{\partial} f(x)} \langle w, y - x\rangle$$

for all $y \in B_\varepsilon(x)$.

At a point x for given $(0, w) \in \overline{d}f(x)$ consider the set

$$L_w(x) = (0, w) + \underline{d}f(x).$$

Proposition 2 *A point $x^* \in \mathbb{R}^n$ is a local minimizer of Problem (3) if and only if the following condition holds:*

$$0_{n+1} \in L_w(x^*) \quad \forall (0, w) \in \overline{d}f(x). \qquad (8)$$

DP Programming: local optimality conditions

At a point x for a given $w \in \overline{\partial} f(x)$ consider the set

$$\mathcal{L}_w(x) = w + \underline{\partial} f(x).$$

Proposition 3 *The condition (8) is equivalent to the following condition:*

$$0_n \in \mathcal{L}_w(x^*) \quad \forall w \in \overline{\partial} f(x^*). \tag{9}$$

Corollary 2 *A point x^* is a local minimizer of Problem (3) if and only if the condition (9) holds.*

Consider the set

$$L_{\theta w}(x) = (\theta, w) + \underline{d}f(x), \quad (\theta, w) \in \overline{d}f(x).$$

Proposition 4 *Suppose that at a point* $x^* \in \mathbb{R}^n$

$$0_{n+1} \in L_{\theta w}(x^*) \quad \forall (\theta, w) \in \overline{d}f(x^*). \tag{10}$$

Then x^* *is a global minimizer of Problem (3).*

For a given point $x \in \mathbb{R}^n$ take any $(\theta, w) \in \overline{d}f(x)$ and define the following sets:

$$\underline{d}_\theta f(x) = \{((\eta, v) \in \underline{d}f(x) : \ \eta + \theta \geq 0\},$$

$$L^+_{\theta w}(x) = (\theta, w) + \underline{d}_\theta f(x).$$

It is clear that $\underline{d}_\theta f(x) \neq \emptyset$ for any $(\theta, w) \in \overline{d}f(x)$.

Proposition 5 *A point $x^* \in \mathbb{R}^n$ is a global minimizer of Problem (3) if and only if $0_n \in \{v : (\eta, v) \in L^+_{\theta w}(x^*)\}$ for any $(\theta, w) \in \overline{d}f(x^*)$.*

DP Programming: numerical algorithms

Assume that $x \in \mathbb{R}^n$ is not an inf-stationary point which means that

$$\partial f_2(x) \not\subset \partial f_1(x).$$

Let

$$R(x) = \{(i,j) \in I_1 \times I_2 : \ i \in R_1(x), \ j \in R_2(x)\}.$$

At the point x we compute $\bar{v} \in \partial f_1(x)$ and $\bar{w} \in \partial f_2(x)$ such that

$$\|\bar{v} - \bar{w}\| = \max_{w \in \partial f_2(x)} \ \min_{v \in \partial f_1(x)} \|v - w\|. \tag{11}$$

It is clear that $\|\bar{v} - \bar{w}\| > 0$. Define the direction

$$\bar{d} = -(\bar{v} - \bar{w}). \tag{12}$$

A direction $d^0 = \|\bar{d}\|^{-1}\bar{d}$ is the steepest descent direction of f at x.

Compute $\bar{\lambda}$ by

$$\bar{\lambda} = \sup\{\lambda : \ R(x + \lambda\bar{d}) \subset R(x)\}.$$

If $\bar{\lambda} = \infty$, then the function f is unbounded along the ray $\{x + \lambda\bar{d} : \lambda \geq 0\}$. Main properties of the direction $\bar{d}$ are summarized in the following proposition.

Proposition 6 *Assume that $x \in \mathbb{R}^n$ is not an inf-stationary point and the direction $\bar{d}$ is defined by (12). Then the following hold:*

- $\bar{\lambda} > 0$;
- $f(x + \lambda\bar{d}) \leq f(x) - \lambda\|\bar{d}\|^2$ *for* $\lambda \in [0, \bar{\lambda})$.

Algorithm 2 Local minimization of DP functions.

Step 1. Select any starting point x^1. Set $k := 1$.

Step 2. If $\partial f_2(x^k) \subset \partial f_1(x^k)$, then stop. x^k is a local minimizer. Otherwise, select $j_k \in R_2(x^k)$ such that

$$0 \neq c^{j_k} \in \text{Argmax} \left\{ \min_{v \in \partial f_1(x^k)} \|v - w\| : \; w \in \partial f_2(x^k) \right\} \tag{13}$$

and solve the following problem:

$$\text{minimize } g_{j_k}(x) = f_1(x) - \langle c^{j_k}, x \rangle + d_{j_k} \text{ subject to } x \in \mathbb{R}^n. \tag{14}$$

If the problem is unbounded, then stop. Otherwise, let x_*^k be a solution and go to Step 3.

Step 3. Set $x^{k+1} := x_*^k, \; k := k + 1$ and go to Step 2.

Algorithm 2 is finite convergent.

Next we design an algorithm for finding global minimizers of DP functions. Let $\sigma_0 > 0$ be a sufficiently small, $\mu > 0$ be a sufficiently large numbers and $t > 1$.

DP Programming: numerical algorithms

Algorithm 3 Global minimization of DP functions.

Step 1. Select a starting point $x^1 \in \mathbb{R}^n$, set $k := 1$ and $\bar{x}^k := x^k$.

Step 2. Apply Algorithm 2 starting from the point $\bar{x}^k$. This algorithm terminates after finite number of iterations and either finds that the problem is unbounded and computes the local minimizer x^{k+1}.

Step 3. If the problem is unbounded then the algorithm terminates. Otherwise set $k := k + 1$ and $\sigma := \sigma_0$.

Step 4. Compute $\bar{z}^k(\sigma) = (\bar{a}_k(\sigma), \bar{v}^k(\sigma)) \in \mathbb{R}^{n+1}$ such that

$$\|\bar{z}^k(\sigma)\|^2 = \max_{(\sigma,w)\in \bar{d}f(x^k)} \min\{\|z\|^2 : z \in L_{w\sigma}(x^k)\}.$$

Step 5. If $\|\bar{z}^k(\sigma)\| > 0$ then set $\bar{x}^k := x^k - \bar{a}_k(\sigma)\bar{v}^k(\sigma)$ and go to Step 2.

Step 6. If $\|\bar{z}^k(\sigma)\| = 0$ then set $\sigma := t\sigma$. If $\sigma > \mu$ then stop, x^k is a global minimizer. Otherwise go to Step 4.

Algorithm 3 is finite convergent.

$$\begin{aligned} &\text{minimize } f(x) \\ &\text{subject to } \langle g_i, x\rangle \le p_i, i = 1 \ldots, l \end{aligned} \tag{15}$$

For a given $x \in \mathbb{R}^n$, let $\alpha_i(x) = \langle g, x\rangle - p_i$.

Proposition 7 *A point $x^* \in \mathbb{R}^n$ is a global minimizer of Problem (15) if and only if $0_n \in$ co$\{v : (\eta, v) \in L^+_{\theta w}(x^*)\} \cup \{(\alpha_i(x^*), g_i)\}$ for any $(\theta, w) \in \overline{d}f(x^*)$.*

Conclusions and Acknowledgements

Australian Research Council (Project number: DP140103213).

THANK YOU!