DC Programming: A brief tutorial.

Andres Munoz Medina
Courant Institute of Mathematical Sciences
munoz@cims.nyu.edu

Difference of Convex Functions

- Definition: A function f is said to be DC if there exists $g,\ h$ convex functions such that $f = g - h$
- Definition: A function is locally DC if for every x there exists a neighborhood U such that $f|_U$ is DC.

Contents

- Motivation.
- DC functions and properties.
- DC programming and DCA.
- CCCP and convergence results.
- Global optimization algorithms.

Motivation

- Not all machine learning problems are convex anymore
- Transductive SVM's [Wang et. al 2003]
- Kernel learning [Argyriou et. al 2004]
- Structure prediction [Narasinham 2012]
- Auction mechanism design [MM and Muñoz]

Notation

- Let g be a convex function. The conjugate of g is defined as g^*

$$g^*(y) = \sup_x \langle y, x\rangle - g(x)$$

- For $\epsilon > 0, \partial_\epsilon g(x_0)$ denotes the ϵ subdifferential of g at x_0 , i.e.

$$\partial_\epsilon g(x_0) = \{v \in \mathbb{R}^n | g(x) \geq g(x_0) + \langle x - x_0, v\rangle - \epsilon$$

- $\partial g(x_0)$ will denote the exact subdifferential.

DC functions

- A function $f : \mathbb{R} \to \mathbb{R}$ is DC iff is the integral of a function of bounded variation. [Hartman 59]
- A locally DC function is globally DC. [Hartman 59]
- All twice continuously differentiable functions are DC.
- Closed under sum, negation, supremum and products.

DC programming

- DC programming refers to optimization problems of the form.

$$\min_x g(x) - h(x)$$

- More generally, for $f_i(x)$ DC functions

$$\min_x \; g(x) - h(x)$$

$$\text{subject to } f_i(x) \leq 0$$

Global optimality conditions.

- A point x^* is a global solution if and only if $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*)$

- Let $w^* = \inf_x g(x) - h(x)$, then a point x^* is a global solution if and only if

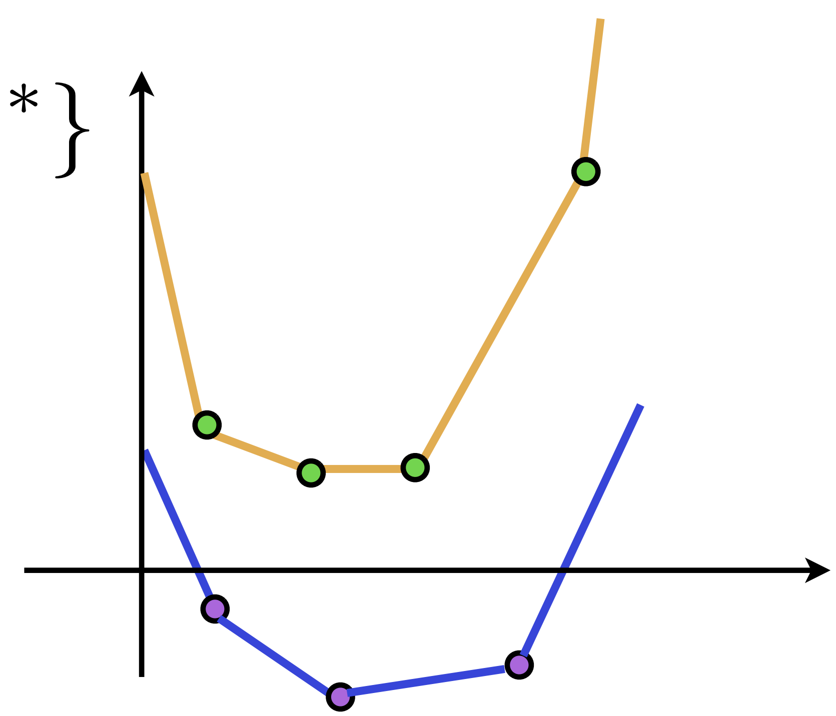
$$0 = \inf_x \{-h(x) + t | g(x) - t \leq w^*\}$$

Global optimality conditions.

- A point x^* is a global solution if and only if $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*)$

- Let $w^* = \inf_x g(x) - h(x)$, then a point x^* is a global solution if and only if

$$0 = \inf_x \{-h(x) + t | g(x) - t \leq w^*\}$$

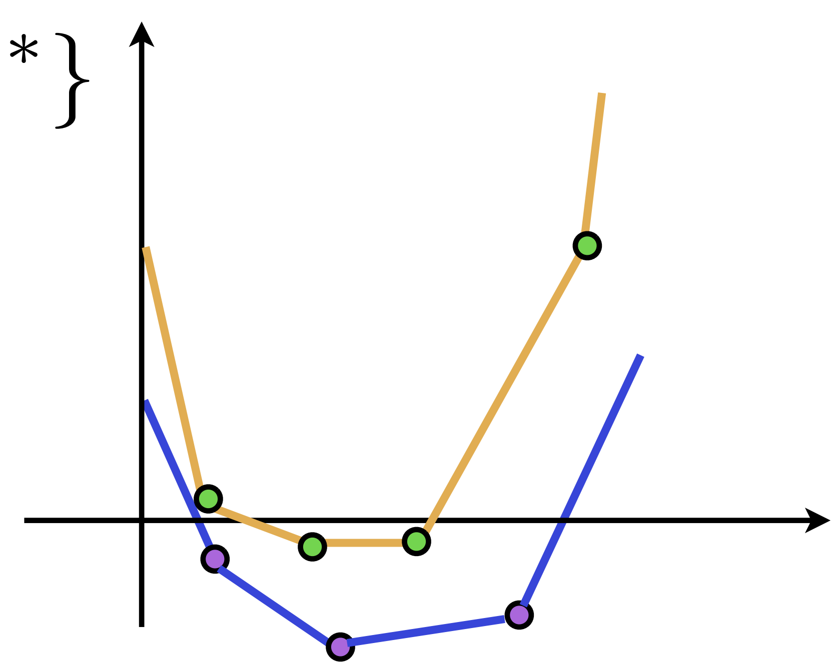


Global optimality conditions.

- A point x^* is a global solution if and only if $\partial_\epsilon h(x^*) \subset \partial_\epsilon g(x^*)$

- Let $w^* = \inf_x g(x) - h(x)$, then a point x^* is a global solution if and only if

$$0 = \inf_x \{-h(x) + t | g(x) - t \leq w^*\}$$



Local optimality conditions

- If x^* verifies $\partial h(x^*) \subset \mathrm{int} \partial g(x^*)$, then is a strict local minimizer of $g - h$

DC duality

- By definition of conjugate function

$$\inf_x g(x) - h(x) = \inf_x g(x) - (\sup_y \langle x, y \rangle - h^*(y))$$
$$= \inf_x \inf_y h^*(y) + g(x) - \langle x, y \rangle$$
$$= \inf_y h^*(y) - g^*(y)$$

- This is the dual of the original problem

DC algorithm.

- We want to find a sequence x_k that decreases the function at every step.
- Use duality. If $y \in \partial h(x_0)$ then

$$
\begin{aligned}
h^*(y) - g^*(y) &= \langle x_0, y\rangle - h(x_0) - \sup_x(\langle x, y\rangle - g(x)) \\
&= \inf_x g(x) - h(x) + \langle x_0 - x, y\rangle \\
&\leq g(x_0) - h(x_0)
\end{aligned}
$$

DC algorithm

- Solve the partial problems

$$S(x^*) \qquad \inf\{h^*(y) - g^*(y) : y \in \partial h(x^*)\}$$

$$T(y^*) \qquad \inf\{g(x) - h(x) : x \in \partial g^*(y^*)\}$$

- Choose $y_k \in S(x_k)$ and $x_{k+1} \in T(y_k)$.
- Solve concave minimization problems.
- Simplified DCA

$$y_k \in \partial h(x_k) \qquad x_k \in \partial g^*(y_k)$$

DCA as CCCP

- If the function is differentiable, the simplified DCA becomes $y_k = \nabla h(x_k)$ and $x_{k+1} \in \partial g^*(y_k)$

- Equivalent to

DCA as CCCP

- If the function is differentiable, the simplified DCA becomes $y_k = \nabla h(x_k)$ and $x_{k+1} \in \partial g^*(y_k)$

  $\rightarrow$

- Equivalent to

DCA as CCCP

- If the function is differentiable, the simplified DCA becomes $y_k = \nabla h(x_k)$ and $x_{k+1} \in \partial g^*(y_k)$

  $\rightarrow$ $x_{k+1} \in \mathrm{argmin}\ g(x) - \langle x, \nabla h(x_k) \rangle$

- Equivalent to

DCA as CCCP

- If the function is differentiable, the simplified DCA becomes $y_k = \nabla h(x_k)$ and $x_{k+1} \in \partial g^*(y_k)$

  ➔ $x_{k+1} \in \text{argmin } g(x) - \langle x, \nabla h(x_k) \rangle$

- Equivalent to

$$x_{k+1} \in \text{argmin } g(x) - h(x_k) - \langle x - x_k, \nabla h(x_k) \rangle$$

CCCP as a majorization minimization algorithm

- To minimize f , MM algorithms build a majorization function F such that

$$f(x) \leq F(x, y) \; \forall x, y$$
$$f(x) = F(x, x) \; \forall x$$

- Do coordinate descent on F
- In our scenario

$$F(x, y) = g(x) - h(y) - \langle x - y, \nabla h(y) \rangle$$

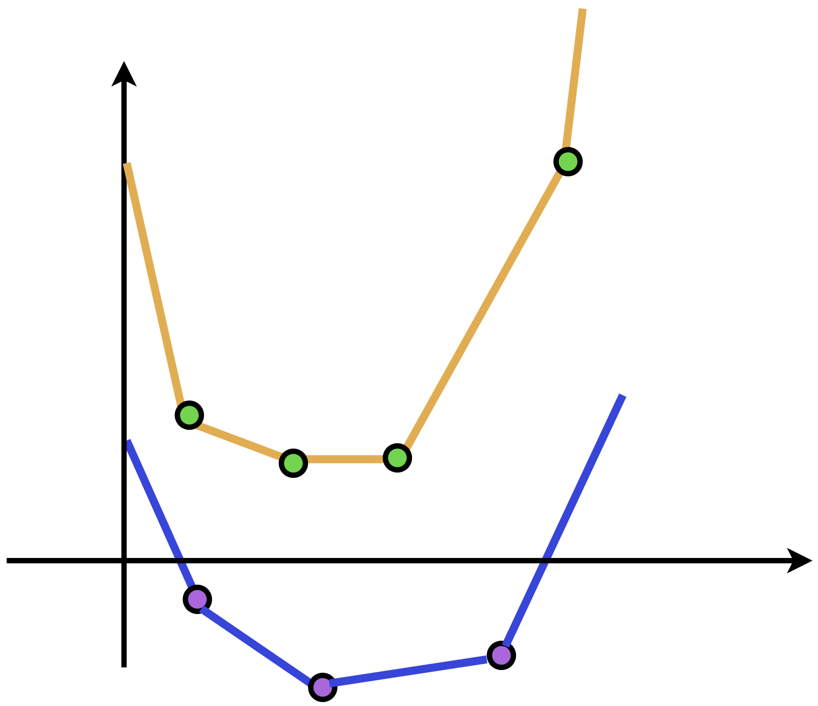
Convergence results

- Unconstrained DC functions: Convergence to a local minimum (no rate of convergence). Bound depends on moduli of convexity. [PD Tao, LT Hoai An 97]
- Unconstrained smooth optimization: Linear or almost quadratic convergence depending on curvature [Roweis et. al 03]
- Constrained smooth optimization: Convergence without rate using Zangwill's theory. [Lanckriet, Sriperumbudur 09]

Global convergence

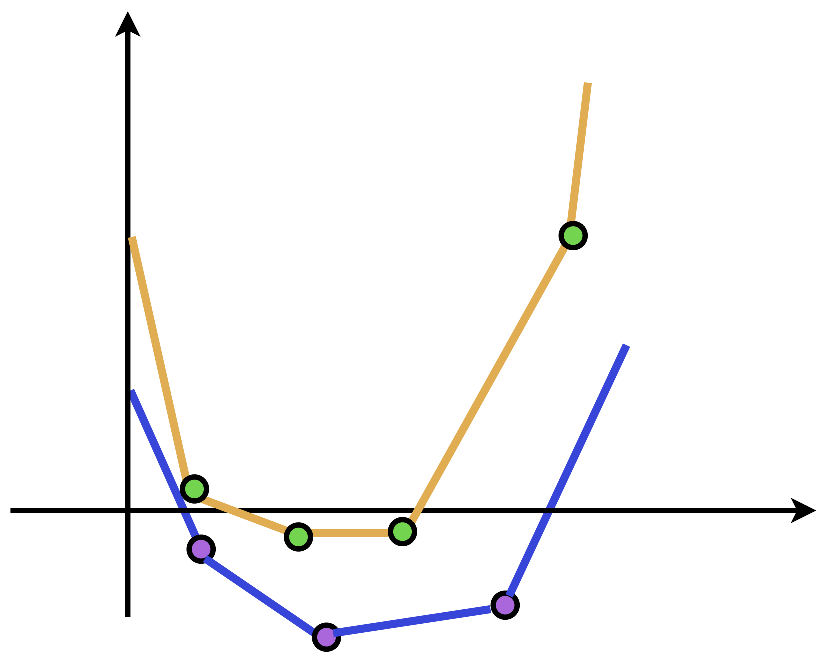
- NP-hard in general: Minimizing a quadratic function with one negative eigenvalue with linear constraints. [Pardalos 91]
- Mostly branch and bound methods and cutting plane methods [H. Tuy 03, Horst and Thoai 99]
- Some successful results with low rank functions.

Cutting plane algorithm[H. Tuy 03]



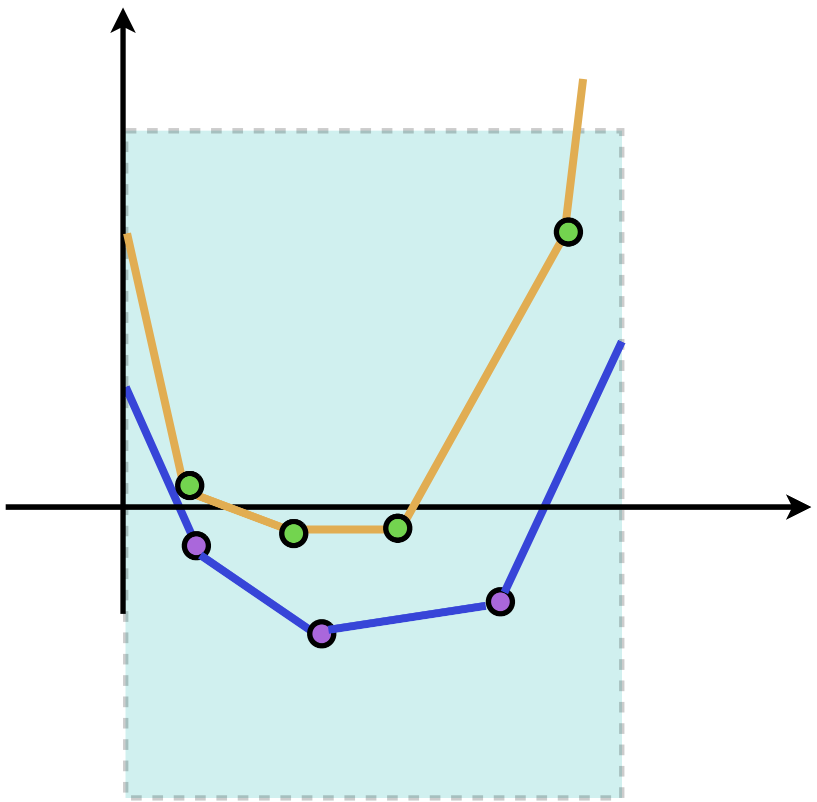
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm [H. Tuy 03]



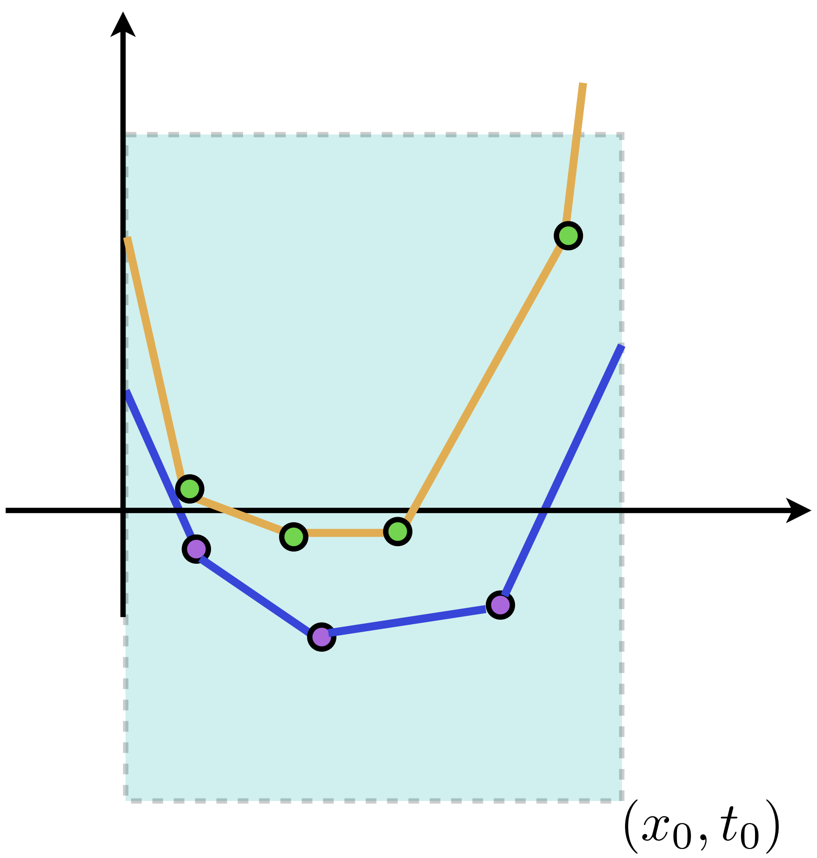
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



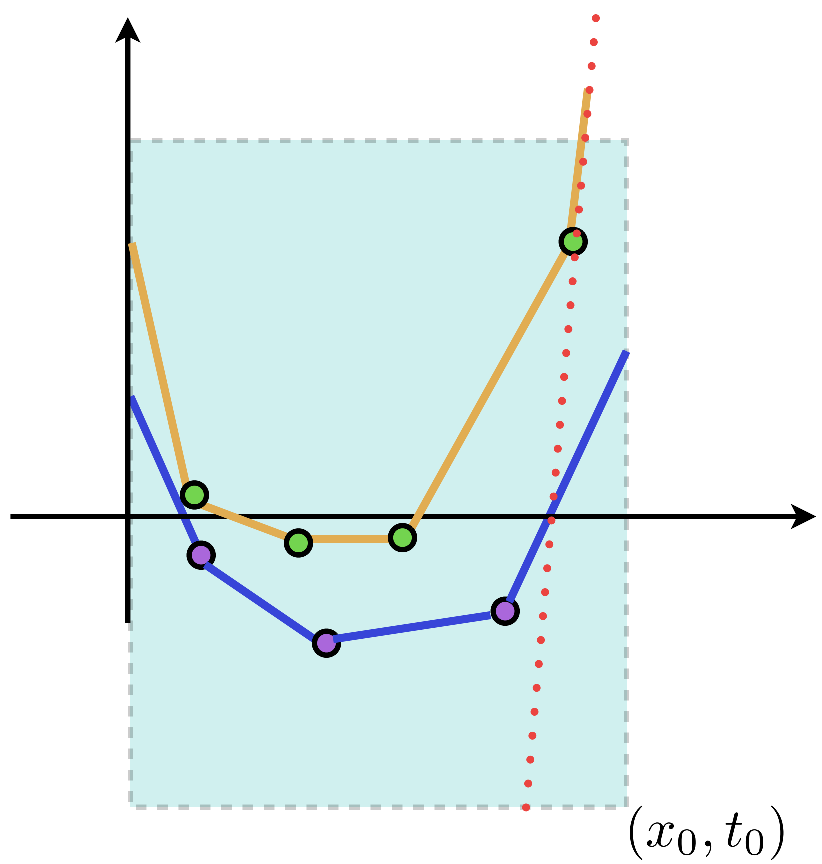
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



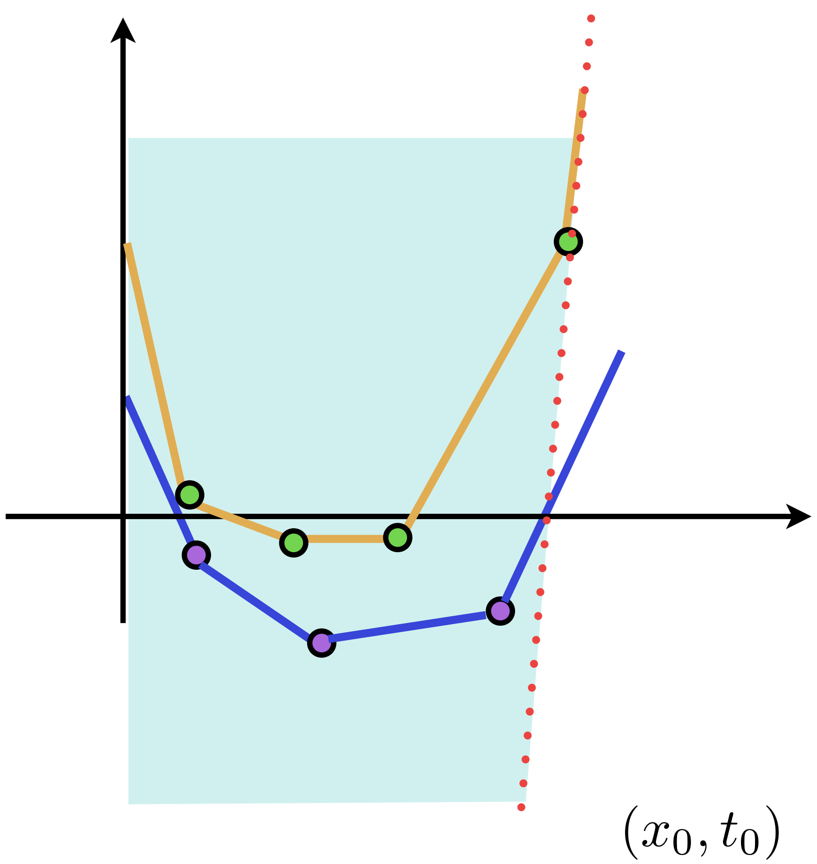
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



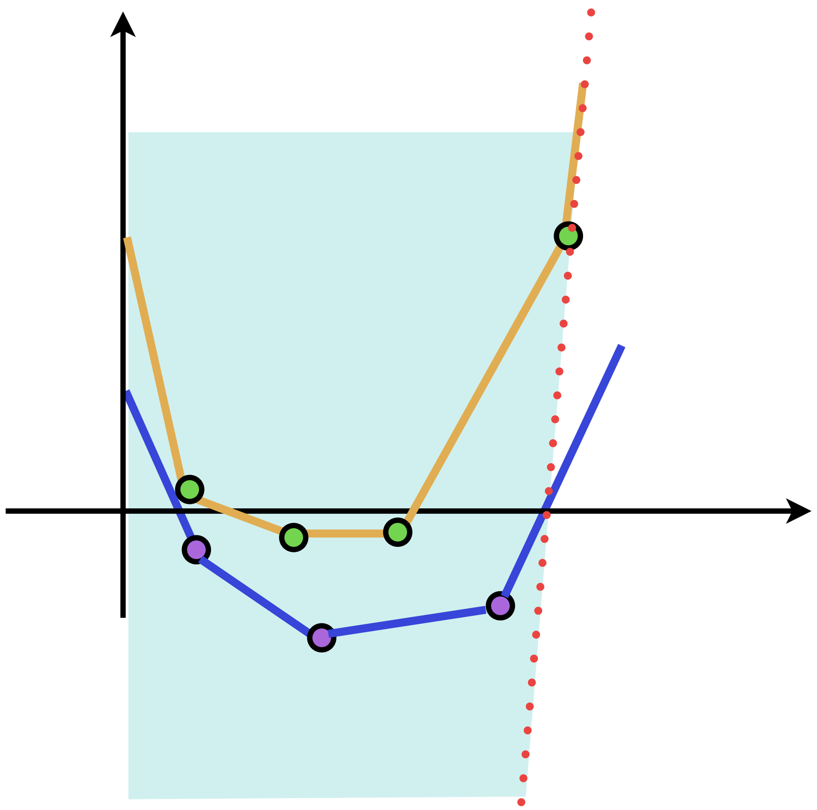
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



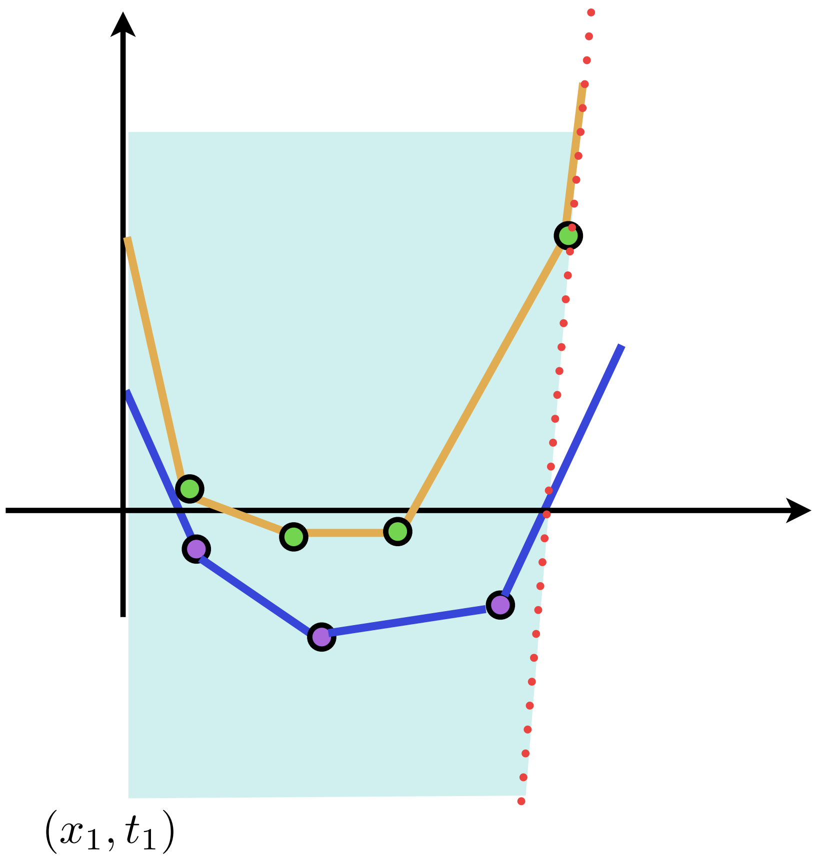
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



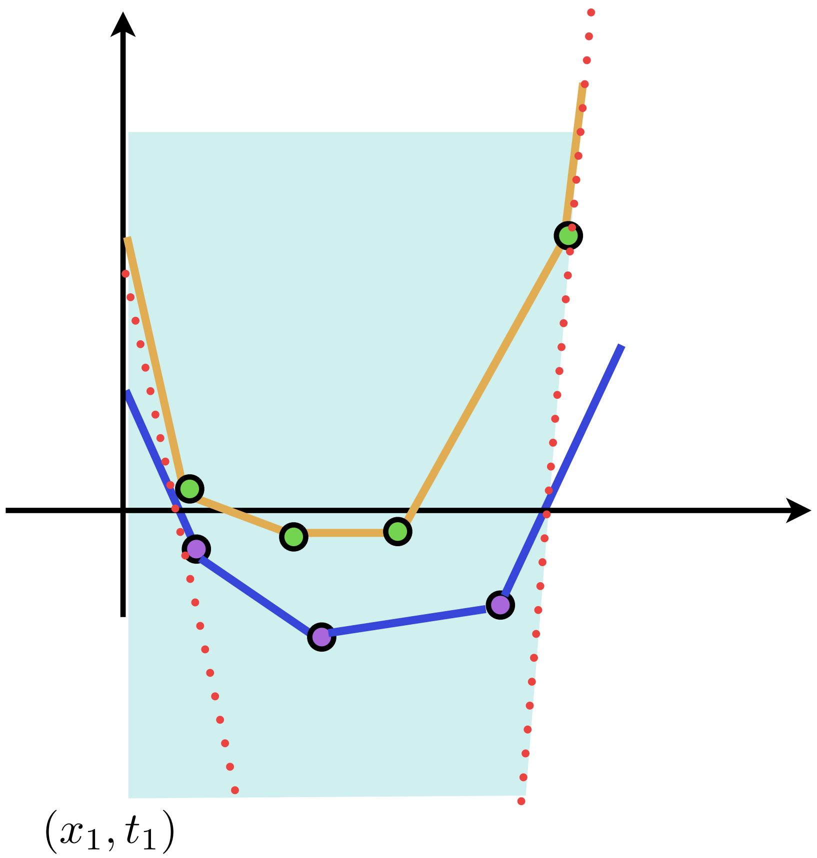
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm [H. Tuy 03]



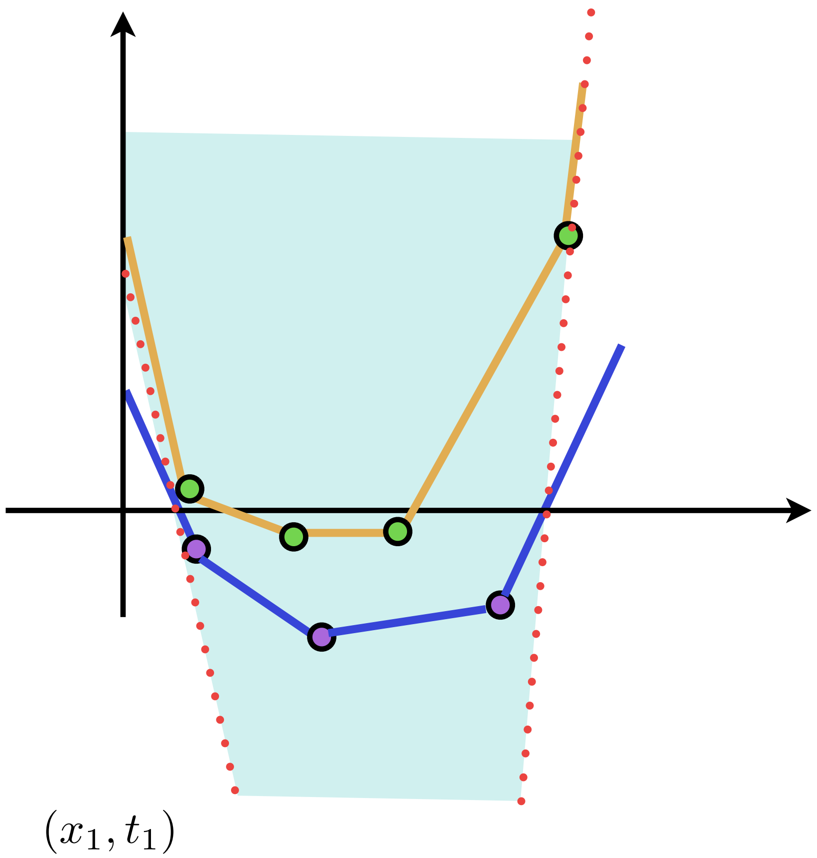
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



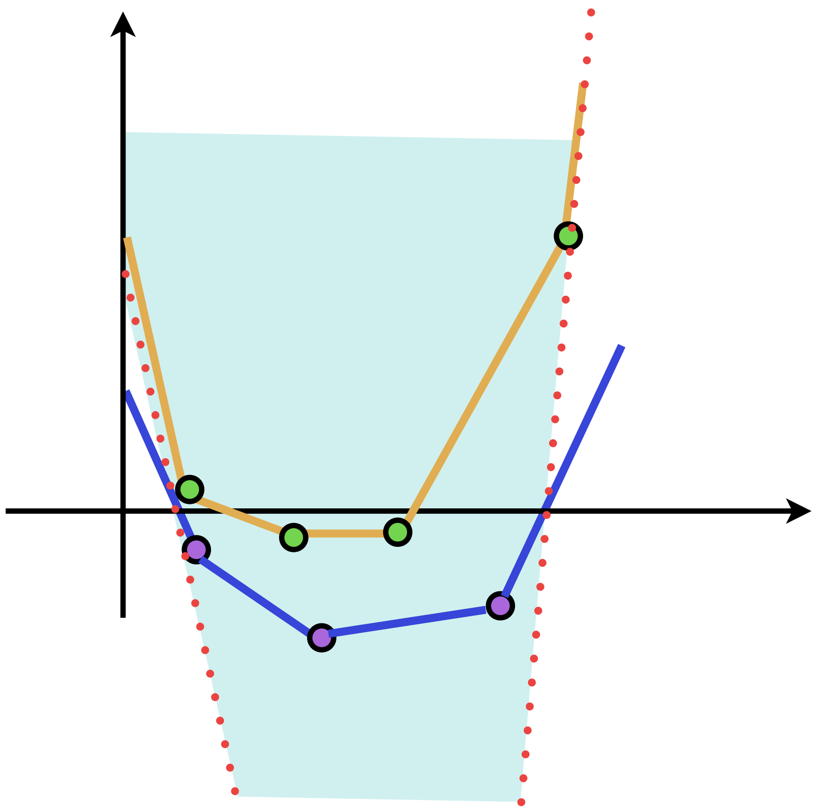
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



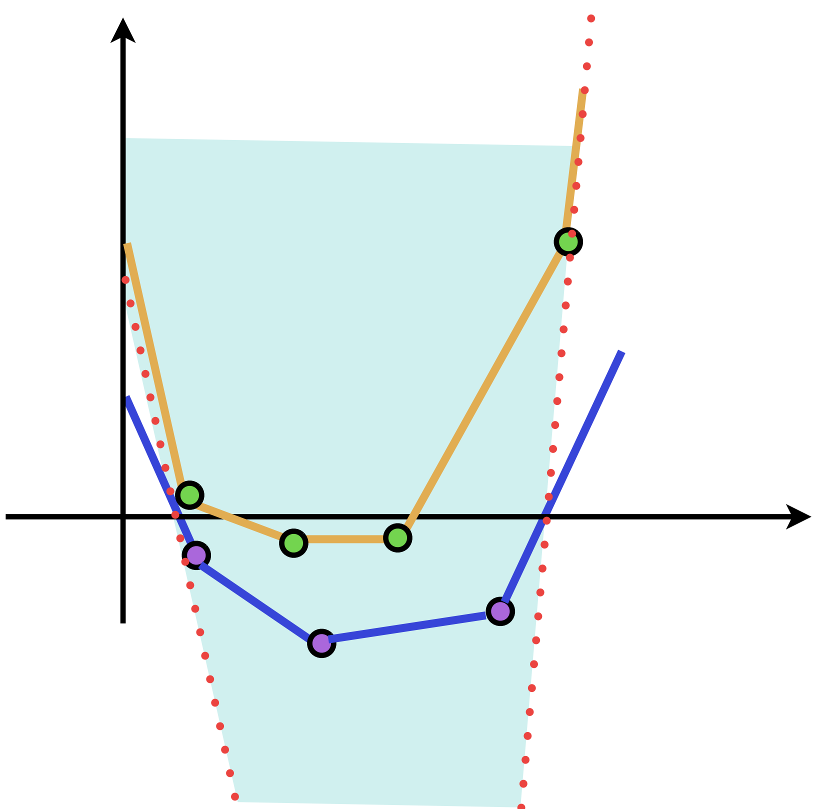
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



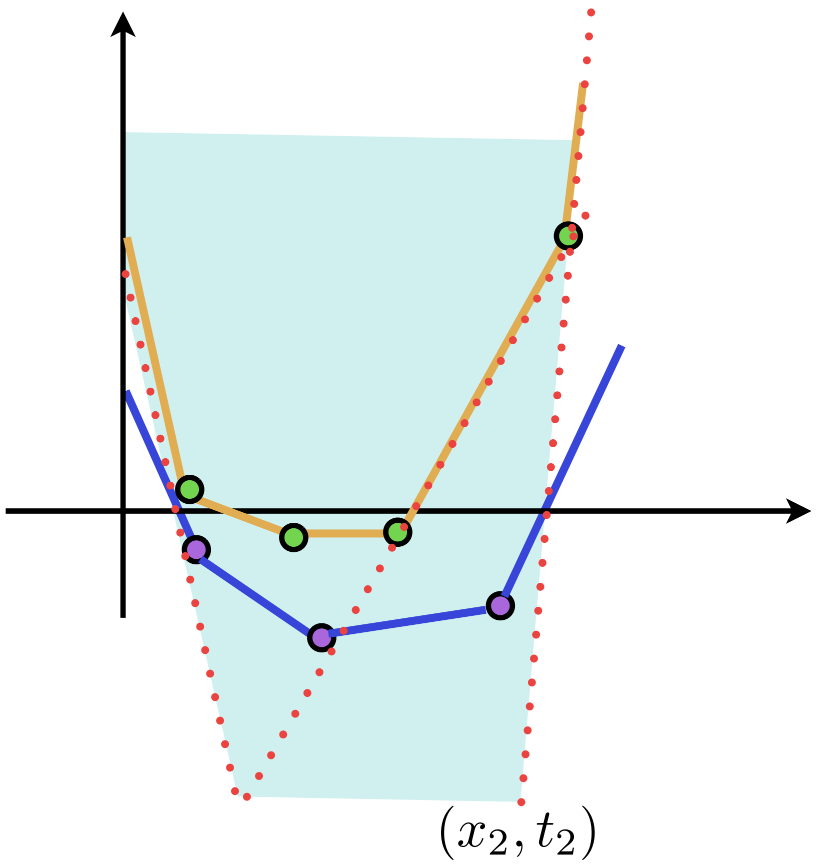
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



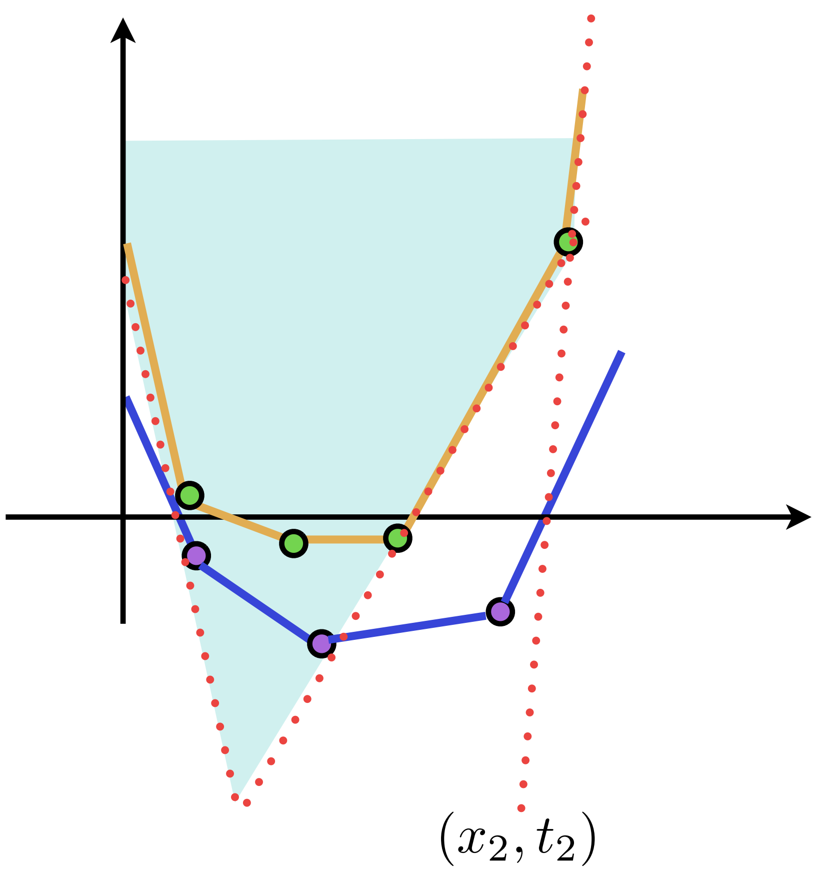
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



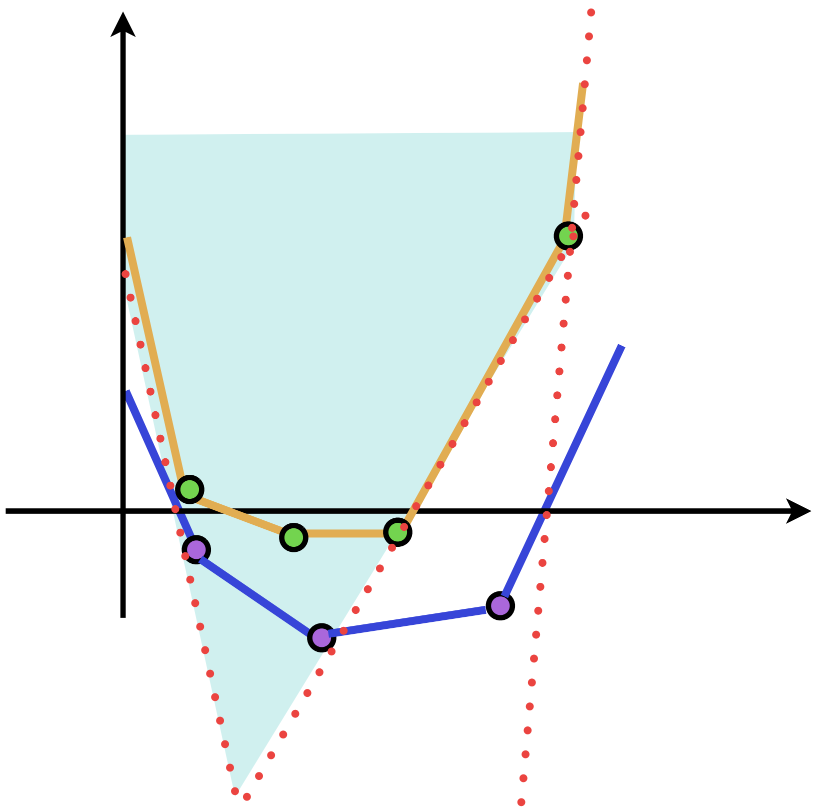
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm [H. Tuy 03]



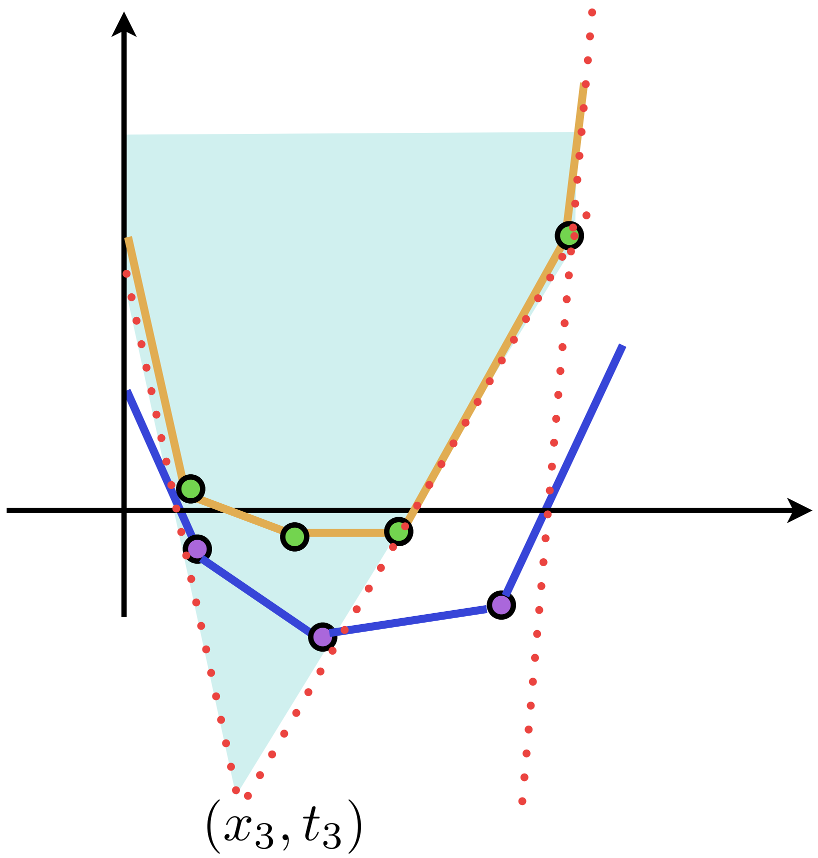
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



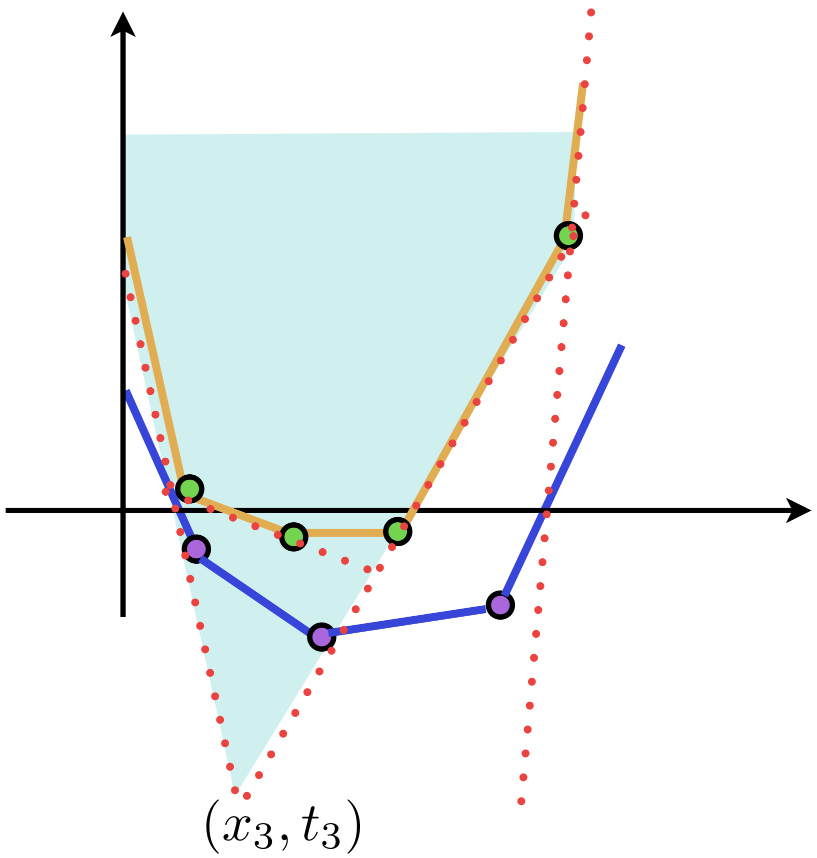
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



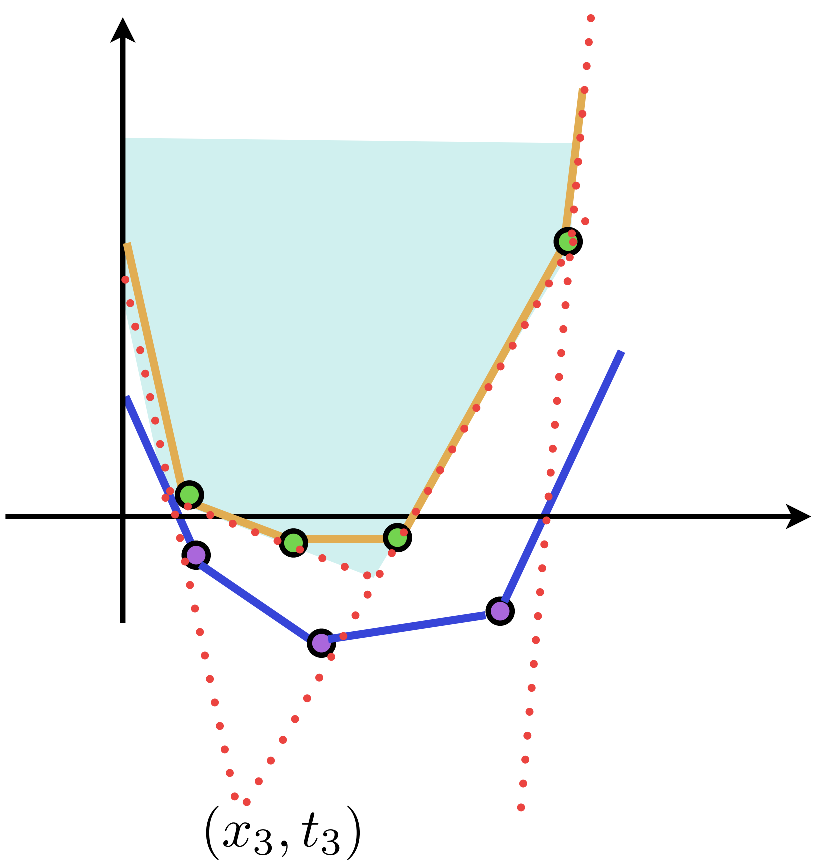
- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Cutting plane algorithm[H. Tuy 03]



- All limit points of this algorithm are global minimizers
- Finite convergence for piecewise linear functions (Conjecture).
- Keeps track of exponentially many vertices.

Branch and bound. [Hoarst and Thoai 99]

- Main idea is to keep upper and lower bounds of $g - h$ on simplices $\mathcal{S}_k$.
- Upper bound: Evaluate $g(x_k) - h(x_k)$ for $x_k \in \mathcal{S}_k$. Use DCA as subroutine for better bounds.
- Lower bound: If $(v_i)_{i=1}^{n+1}$ are the vertices of $\mathcal{S}_k$ and $x = \sum_{i=1}^{n+1} \alpha_i v_i$. Solve

$$\min_{\alpha} g\Big(\sum \alpha_i v_i \Big) - \sum \alpha_i h(v_i)$$

Low rank optimization and polynomial guarantees[Goyal and Ravi 08]

- A function $f : \mathbb{R}^n \to \mathbb{R}$ has rank $k \ll n$ if there exists $g : \mathbb{R}^k \to \mathbb{R}$ and $\alpha_1, \ldots, \alpha_k \in \mathbb{R}^n$ such that $f(x) = g(\alpha_1 \cdot x, \ldots, \alpha_k \cdot x)$
- Most examples in Economy literature.
- For a quasi-concave function f we want to solve $\min_{x \in C} f(x)$.
- Can always transform DC programs to this type of problem.

Algorithm

- Let g satisfy the following conditions.
  - The gradient $\nabla g(y) \geq 0$
  - $g(\lambda y) \leq \lambda^c g(y)$ for all $\lambda > 1$ and some c
  - $\alpha_i \cdot x > 0$ for all $x \in P$
- There is an algorithm that finds $\widetilde{x}$ with $f(\widetilde{x}) \leq (1+\epsilon) f(x^*)$ in $O\left(\frac{c^k}{\epsilon^k}\right)$

Further reading

- Farkas type results and duality for DC programs with convex constraints. [Dinh et. al 13]
- On DC functions and mappings [Duda et. al 01]

Open problems

- Local rate of convergence for constrained DC programs.
- Is there a condition under which DCA finds global optima. For instance $g - h$ might not be convex but $h^* - g^*$ might.
- Finite convergence of cutting plane methods.

References

- Dinh, Nguyen, Guy Vallet, and T. T. A. Nghia. "Farkas-type results and duality for DC programs with convex constraints." (2006).
- Goyal, Vineet, and R. Ravi. "An FPTAS for minimizing a class of low-rank quasi-concave functions over a convex set." *Operations Research Letters* 41.2 (2013): 191-196.
- Tao, Pham Dinh, and Le Thi Hoai An. "Convex analysis approach to DC programming: theory, algorithms and applications." *Acta Mathematica Vietnamica* 22.1 (1997): 289-355.
- Horst, R., and Ng V. Thoai. "DC programming: overview." *Journal of Optimization Theory and Applications* 103.1 (1999): 1-43.
- Tuy, H. "On global optimality conditions and cutting plane algorithms." *Journal of optimization theory and applications* 118.1 (2003): 201-216.
- Yuille, Alan L., and Anand Rangarajan. "The concave-convex procedure."*Neural Computation* 15.4 (2003): 915-936.

References

- Salakhutdinov, Ruslan, Sam Roweis, and Zoubin Ghahramani. "On the convergence of bound optimization algorithms." *Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers Inc., 2002.
- Hartman, Philip. "On functions representable as a difference of convex functions." *Pacific J. Math* 9.3 (1959): 707-713.
- Hiriart-Urruty, J-B. "Generalized Differentiability/Duality and Optimization for Problems Dealing with Differences of Convex Functions." *Convexity and duality in optimization*. Springer Berlin Heidelberg, 1985. 37-70.