Global Piecewise Linear Optimization Algorithm for Data-driven Modeling

Shuning Wang

Department of Automation, Tsinghua University

July 14, 2014

Outline

1. Brief Introduction to Piecewise Linear Functions
2. Concave Optimization Problems in DDM
3. Hill-detouring Approach
4. Extension of Hill-detouring Approach
5. Conclusions

1. Brief Introduction to Piecewise Linear Function

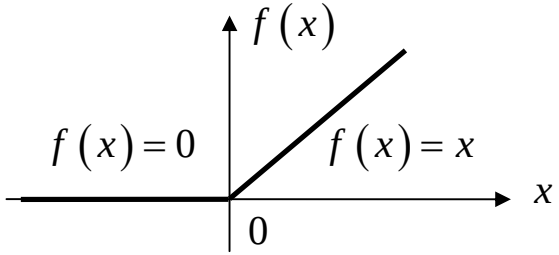
Basic definition

$f : R^n \mapsto R$ **is a piecewise linear function (PLF) defined on a set** $D \subseteq R^n$ **if** D **can be partitioned into mutually disjoint subsets** $D_i \subseteq D, i = 1,2,\cdots,m$ **and there are** $a_i \in R^n,\ b_i \in R,\ i = 1,2,\cdots,m$ **such that**

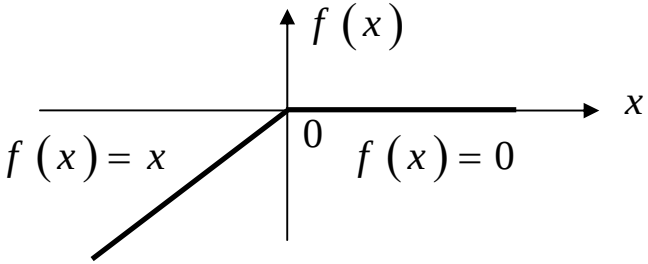
$$f(x) = a_i^T x + b_i,\ \forall x \in D_i,\ i = 1,2,\cdots,m$$

Examples

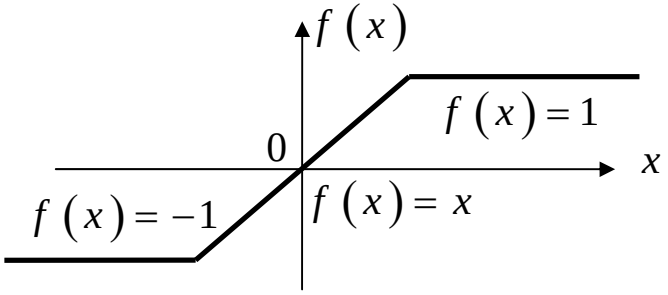
$$f(x)=\begin{cases}0 & \forall x \le 0\\ x & \forall x \ge 0\end{cases}$$



$$f(x)=\begin{cases}x & \forall x \le 0\\ 0 & \forall x \ge 0\end{cases}$$

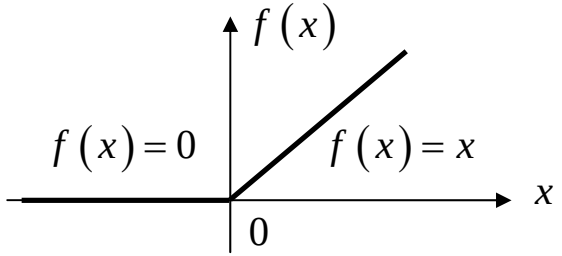


$$f(x)=\begin{cases}-1 & \forall x \le -1\\ x & \forall -1 \le x \le 1\\ 1 & \forall x \ge 1\end{cases}$$

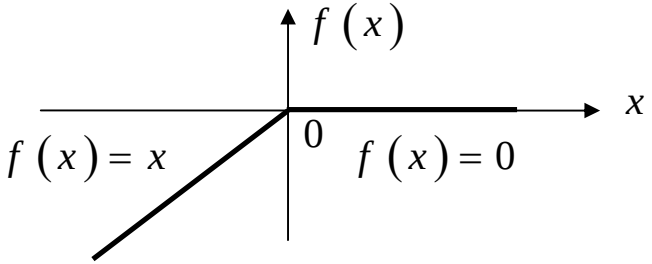


Compact expression

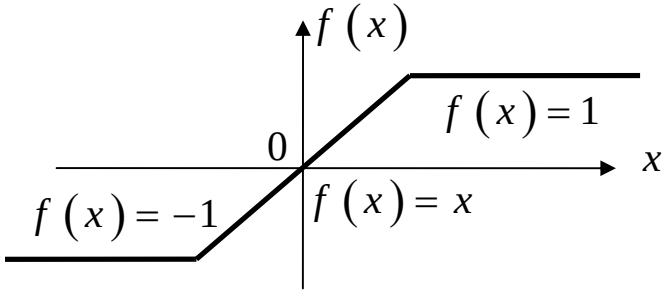
$$f(x) = \max\{0, x\}$$



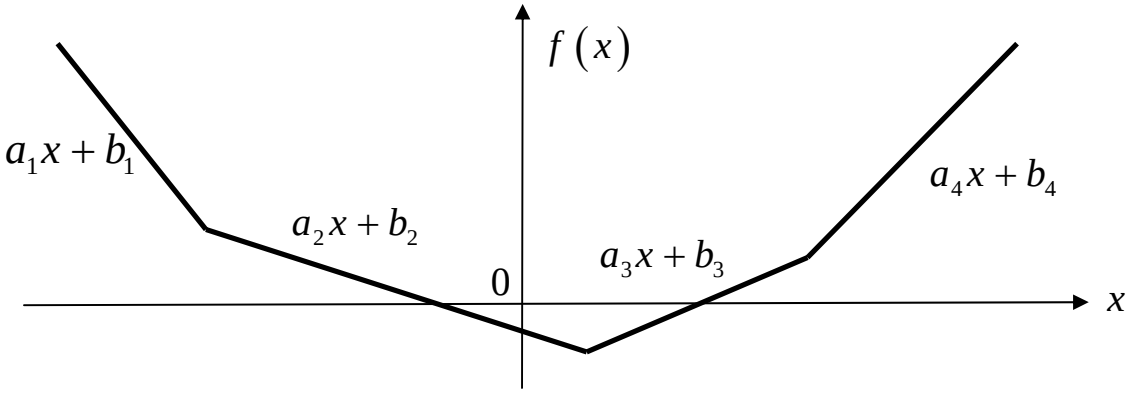
$$f(x) = \min\{0, x\}$$



$$f(x) = \min\{1, \max\{-1, x\}\}$$

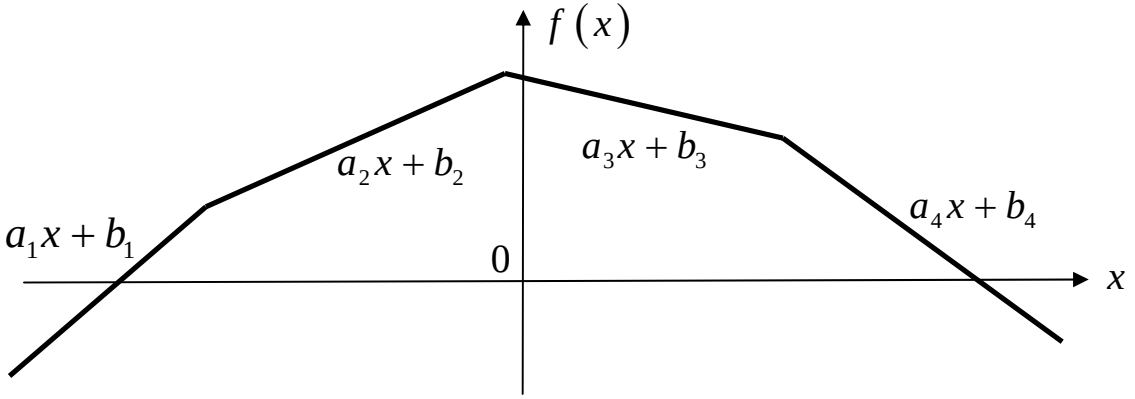


Convex PLF



$$\max_{i=1,2,3,4} a_i x + b_i$$

Concave PLF



$$\min_{i=1,2,3,4} a_i x + b_i$$

Multi dimensional examples

$$\max\left\{0, a^T x + b\right\} \qquad \min\left\{0, a^T x + b\right\} \qquad \min\left\{1, \max\left\{-1, a^T x + b\right\}\right\}$$

$$\max_{i=1,2,\cdots,m} a_i^T x + b_i \qquad \min_{i=1,2,\cdots,m} a_i^T x + b_i \qquad \min_{i=1,2,\cdots,m} \max\left\{0, a_i^T x + b_i\right\}$$

General results

If $f(x), g(x)$ **are PLFs, then** $f(x)+g(x)$, $\max\left\{f(x), g(x)\right\}$, **and** $\min\left\{f(x), g(x)\right\}$ **are PLEs;**

If $f(x), g(x)$ **are convex PLFs, then** $f(x)+g(x)$ **and** $\max\left\{f(x), g(x)\right\}$ **are convex PLFs;**

If $f(x), g(x)$ **are concave PLFs, then** $f(x)+g(x)$ **and** $\min\left\{f(x), g(x)\right\}$ **are concave PLFs**

2. Concave Optimization Problems in DDM

2.1. Basic Optimization Problems in Regression

Given N **pairs data** $y_t \in R,\ x_t \in R^n$ **and** m **basis functions** $\varphi_j : R^n \mapsto R$ **, find** $\theta_j,\ j = 0,1,\cdots,m$ **minimizing fitting errors** $|e_t|, t = 1,2,\cdots N$ **, where** $e_t = y_t - \left(\theta_0 + \sum_{j=1}^{m} \varphi_j (x_t)\theta_j \right)$ **, equivalently**

$$\underset{u_t,\,\theta_j}{\text{minimize}} \quad \{u_t, t = 1,2,\cdots N\}$$

$$\text{subject to} \quad -u_t \le y_t - \left(\theta_0 + \sum_{j=1}^{m} \varphi_j (x_t)\theta_j \right) \le u_t,\ t = 1,2,\cdots,N$$

Using ℓ_1, ℓ_2 or ℓ_∞ norm, the parameters can be found by solving, respectively, the following convex optimization problems

$$\min_{(u,\theta)\in D} \sum_{t=1}^{N} u_t \quad \text{or} \quad \sum_{t=1}^{N} u_t^{\,2} \quad \text{or} \quad \max_{t=1,2,\cdots,N} u_t$$

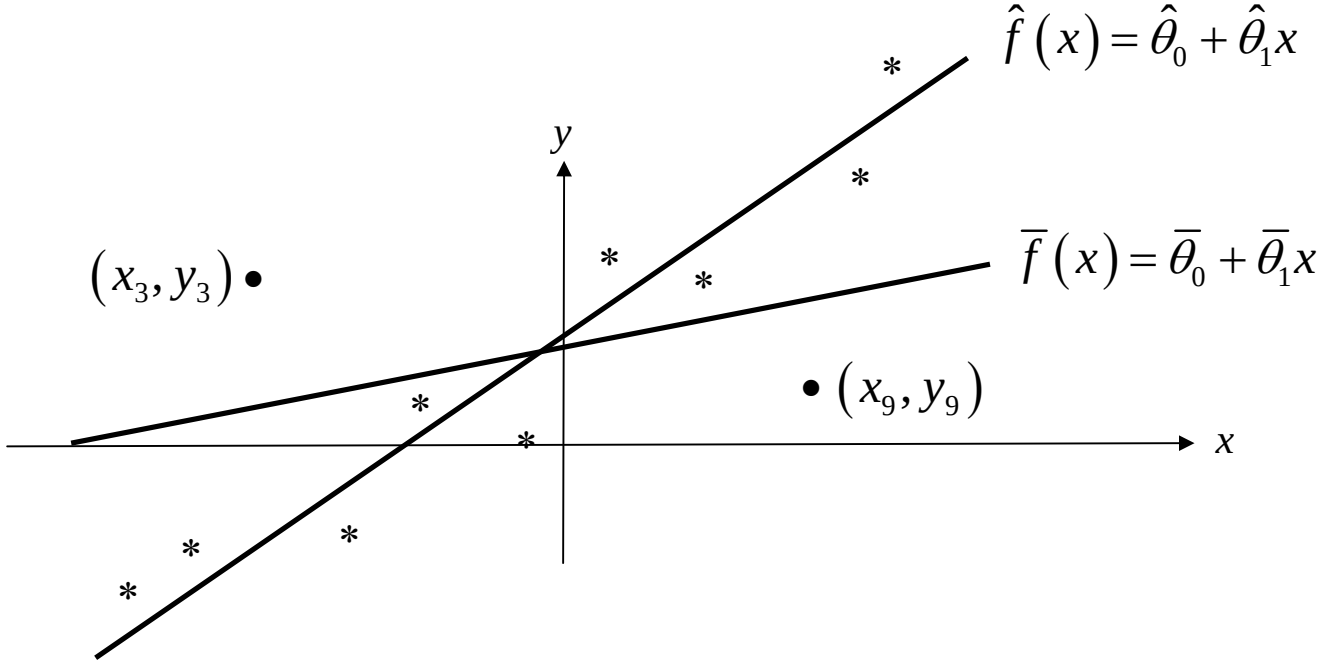
where $u = (u_1, u_2, \cdots, u_N)^T$, $\theta = (\theta_0, \theta_1, \cdots, \theta_m)^T$

$$D = \left\{ (u,\theta) \middle| -u_t \le y_t - \left(\theta_0 + \sum_{j=1}^{m} \varphi_j(x_t)\theta_j \right) \le u_t, \ t = 1, 2, \cdots, N \right\}$$

Here D **is a polyhedron in** R^{N+m+1}

2.2. Robust Estimation (Eliminating Outliers)

Effect of outliers



$\hat{f}$: **Regression result neglecting outliers** $(x_3, y_3), (x_9, y_9)$

$\bar{f}$: **Regression result using all data**

Eliminating outliers by setting a number k

Neglect data whose error is larger than the k-th one, i.e.

$$\underset{(u,\theta)\in D}{\text{minimize}} \quad \left\{ u_{\eta(u,j)},\ j = k, k+1, \cdots, N \right\}$$

where $\eta(u, j)$ denotes the subscript of the j-th large component of $u \in R^N$, which means

$$u_{\eta(u,1)} \geq u_{\eta(u,2)} \geq \cdots \geq u_{\eta(u,N)},\ \ \forall u \in R^N$$

Using ℓ_∞ norm, the robust estimation problem becomes

$$\min_{(u,\theta)\in D} \; u_{\eta(u,k)}$$

When $k = \frac{N}{2}$, it is classical "Least Median of Squares" regression problem (Rousseeuw, 1984).

As $u_{\eta(u,k)}$ is not a concave function (nor a convex one), the above problem is not a concave optimization problem. However, it can be transformed to a concave one, as is shown below.

Define a convex function $f_\tau : R^N \mapsto R$ **for any** $\tau \in \{1, 2, \cdots, N\}$

$$f_\tau(u) = \max\left\{u_{t_1} + u_{t_2} + \cdots + u_{t_\tau},\ \{t_1, t_2, \cdots, t_\tau\} \subseteq \{1, 2, \cdots, N\}\right\}$$

Then $u_{\eta(u,k)}$ **can be written as difference of two convex funtions**

$$\begin{aligned} u_{\eta(u,k)} &= \left(u_{\eta(u,1)} + \cdots + u_{\eta(u,k-1)} + u_{\eta(u,k)}\right) - \left(u_{\eta(u,1)} + \cdots + u_{\eta(u,k-1)}\right) \\ &= f_k(u) - f_{k-1}(u) \end{aligned}$$

The above problem is equivalent to the following concave PL one

$$\min\left\{\lambda - f_{k-1}(u) \,\middle|\, \text{s.t. } f_k(u) \le \lambda, (u, \theta) \in D\right\}$$

Using ℓ_1 norm (to our knowledge, it is not studied in literature), the robust estimation problem becomes

$$\min_{(u,\theta)\in D} \sum_{i=k}^{N} u_{\eta(u,i)}$$

Because

$$\sum_{i=k}^{N} u_{\eta(u,i)} = \min\left\{ \sum_{j=1}^{N-k+1} u_{t_j},\ \{t_1, t_2, \cdots, t_{N-k+1}\} \subseteq \{1, 2, \cdots, N\} \right\}$$

it is a cancave PLF. The above problem is already a concave PL optimization problem.

Remark: ℓ_1 seems better than ℓ_∞ in robust estimation.

Eliminating outliers by setting a threshold $\bar{u}$

Neglect data whose fitting error is not less than $\bar{u}$, i.e.

$$\underset{(u,\theta)\in D}{\text{minimize}} \quad \left\{ u_t, t = 1,2,\cdots,N \,\middle|\, \text{s.t. } u_t < \bar{u} \right\}$$

which is equivalent to

$$\underset{(u,\theta)\in D}{\text{minimize}} \quad \left\{ \min\left\{0, u_t - \bar{u}\right\}, t = 1,2,\cdots,N \right\}$$

Using ℓ_1 norm, the problem becomes concave PL optimization

$$\min_{(u,\theta)\in D} \quad \sum_{t=1}^{N} \min\left\{0, u_t - \bar{u}\right\}$$

2.3. Selection of Basis Functions or Model Reduction

The problem of selecting basis functions can be expressed as

$$\min \ \|v\|_0$$

$$\text{s.t.} \ -\varepsilon \le y_t - \left(\theta_0 + \sum_{j=1}^{m} \varphi_j(x_t)\theta_j \right) \le \varepsilon, \ t = 1, 2, \cdots, N$$

$$-v_j \le \theta_j \le v_j, \ j = 1, 2, \cdots, m$$

where ε is a given threshold of acceptable fitting errors, $\|v\|_0$ is ℓ_0 norm of $v = (v_1, v_2, \cdots, v_m)^T$, which denotes the number of nonzero componets of v

Selecting basis functions by setting a number k

Neglect parameters whose v_i is larger than the k-th one, i.e.

$$\min \sum_{j=k}^{m} v_{\mu(v,j)}$$

$$\text{s.t.} \quad -\varepsilon \le y_t - \left(\theta_0 + \sum_{j=1}^{m} \varphi_j(x_t)\theta_j \right) \le \varepsilon, \quad t = 1, 2, \cdots, N$$

$$-v_j \le \theta_j \le v_j, \quad j = 1, 2, \cdots, m$$

where $v_{\mu(v,1)} \ge v_{\mu(v,2)} \ge \cdots \ge v_{\mu(v,m)}, \quad \forall v = (v_1, v_2, \cdots, v_m)^T \in R^m$

As is explained before, it is a concave PL optimization problem

Selecting basis functions by setting a threshold $\bar{v}$

Neglect parameters whose v_i is not less than $\bar{v}$, i.e.

$$\min \quad \sum_{i=1}^{m} \min\{0, v_i - \bar{v}\}$$

$$\text{s.t.} \quad -\varepsilon \le y_t - \left(\theta_0 + \sum_{j=1}^{m} \varphi_j(x_t)\theta_j \right) \le \varepsilon, \quad t = 1, 2, \cdots, N$$

$$-v_j \le \theta_j \le v_j, \quad j = 1, 2, \cdots, m$$

It is a concave PL optimization problem

3. Hill-detouring Approach (Huang, et al, 2012)

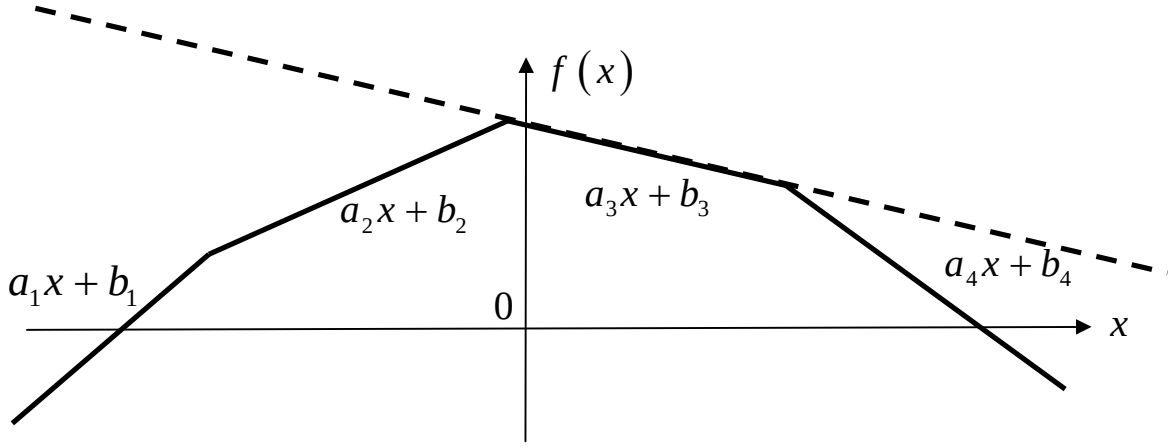
3.1. Two Properties of Cancave PLFs

For any cancave PLF $f : R^n \mapsto R$,

i. If $f(x) = \alpha^T x + \beta,\ \forall x \in D$ **holds for some open set** $D \subset R^n$, **then** $f(x) \le \alpha^T x + \beta,\ \forall x \in R^n$;

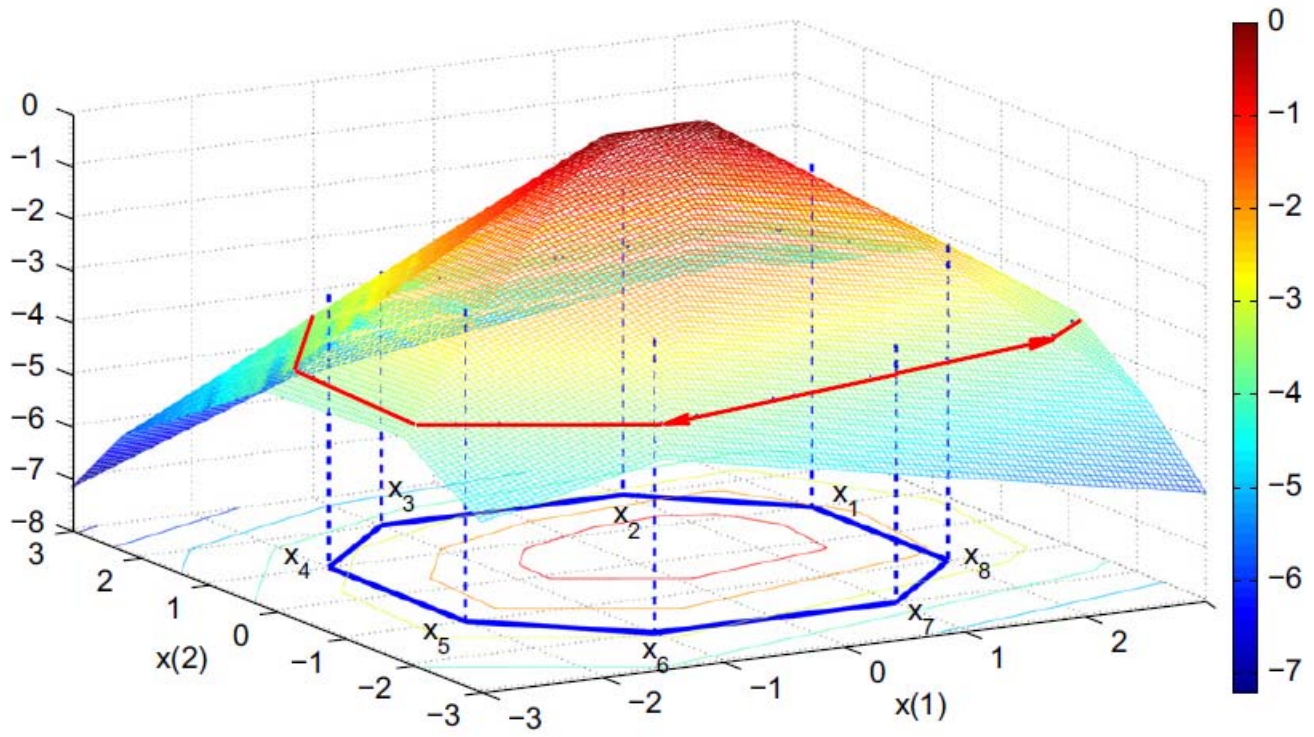
ii. The super level set $\left\{ x \in R^n \middle| f(x) \ge f(\hat{x}) \right\}$ **for any** $\hat{x} \in R^n$ **is a polyhedron**

Illustration of the first property



$$f\left(x\right) \le a_i x + b_i, \;\; \forall x \in R, i = 1,2,3,4$$

Illustration of the second property



Use of the first property

Consider the PL optimization problem

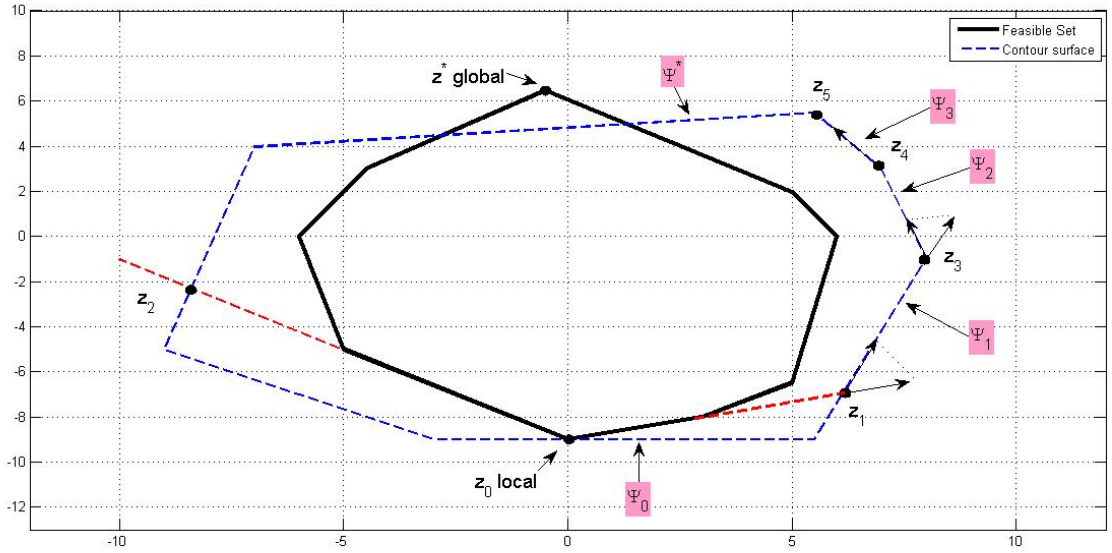
$$\min\left\{ f_1(x) \,\middle|\, \text{s.t. } f_2(x) \le c,\ x \in D \right\} \qquad (1)$$

where $f_1,\ f_2$ **are concave PLFs,** D **is a polyhedron. Given a feasible solution** $\hat{x}$**. Assume that: i.** $f_1(x) = a_1^T x + b_1,\ \forall x \in \hat{D}$ **and** $f_2(x) = a_2^T x + b_2,\ \forall x \in \hat{D}$ **hold for some open set** $\hat{D}$ **containing** $\hat{x}$**; ii.** $\overline{x}$ **is an optimal solution of linear programming problem** $\min\left\{ a_1^T x + b_1 \,\middle|\, \text{s.t. } a_2^T x + b_2 \le c,\ x \in D \right\}$**, then** $f_2(\overline{x}) \le c$**,** $\overline{x} \in D$ **and** $f_1(\overline{x}) \le f_1(\hat{x})$**, i.e.** $\overline{x}$ **is a feasible solution of (1) not worse than** $\hat{x}$

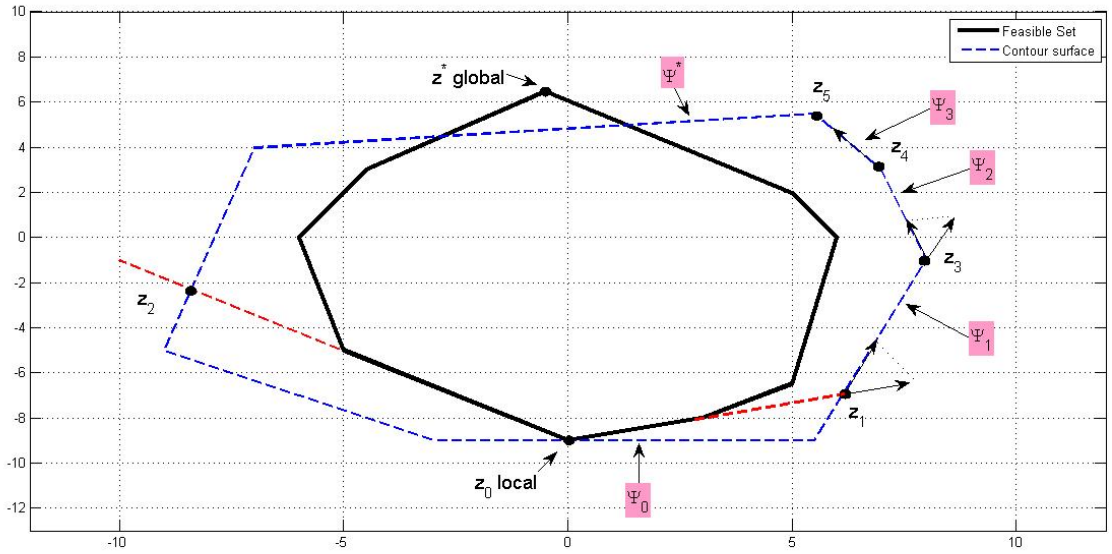
Remark: a locally optimal solution can be obtained using linear programming algorithm

Use of the second property

The contour surface $\left\{ x \in R^n \,\middle|\, f(x) = f(v_0) \right\}$ for any locally optimal solution v_0 consists of finite hyperplanes. Searching them may find a feasible solution $x \neq v_0$ but $f(x) = f(v_0)$, hence escape the local minmum. The following figure illustrates this hill-detouring idea



3.2. Principal Steps in Hill-detouring approach



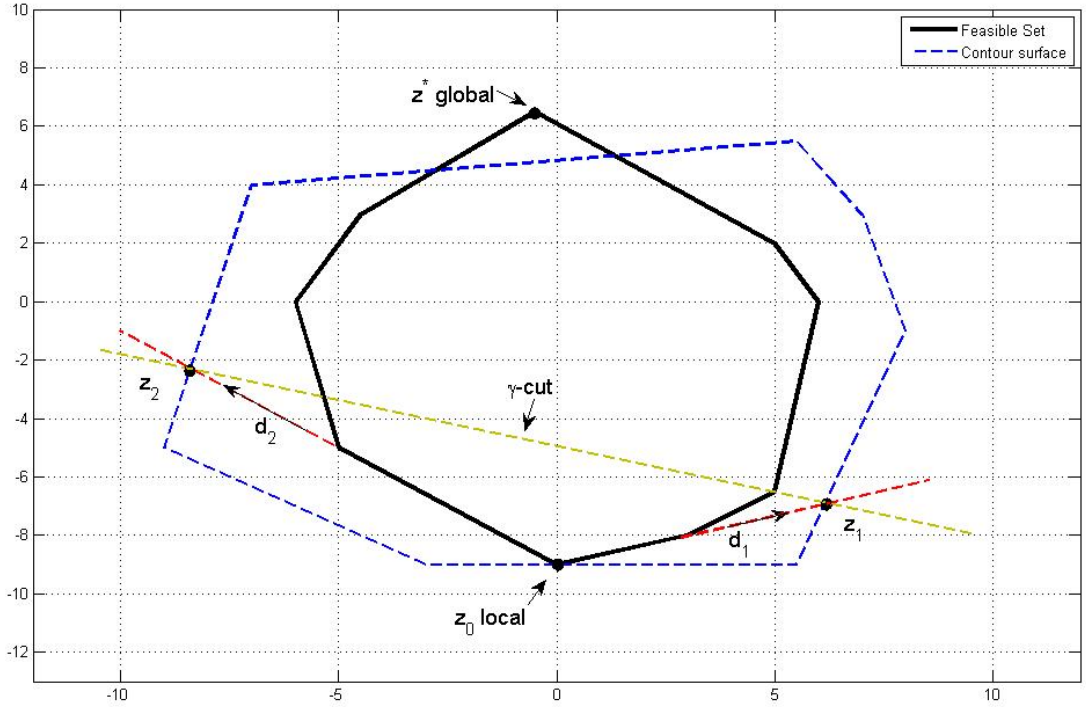
i. Starting from locally optimal point z_0, carry out line search along each edge to find adjacent points(z_i) on contour surface

ii. Choosing point z_1 closest to feasible set, project direction $z_1 - z_0$ on facet Ψ_1, then carry out line search to find z_3

Main shortcoming: no measures are adopted for convergence

4. Extension of Hill-detouring Approach

4.1. Eege directions, γ - extension and γ - cut



The feasible points and contour surface under the γ - cut will not be considered in later optimization or hill-detouring process

Computation of Eege directions, γ - extension and γ - cut

Let $z_0 \in R^n$ be locally optimal (nondegenate) vertex of the cancave optimization problem

$$\min\left\{ f(x) \middle| \text{ s.t. } a_i^T x \le b_i, i = 1,2,\cdots,m \right\}$$

There are $B \in R^{n\times n}$ and $\overline{b} \in R^n$ such that $z_0 = B^{-1}\overline{b}$, and B^{-1} is available using simplex algorithm to get z_0. Let $B^{-1} = \left(d_1\ d_2 \cdots d_n\right)$. Then $d_1, d_2, \cdots, d_n$ are edge directions from z_0 to adjacent vertexes, each $z_i = z_0 + t_i d_i$ with $f(z_i) = f(z_0)$ is a γ - extension, and the hyperplane $\alpha_0^T (x - z_0) = 1$ with $\alpha_0^T = \left(t_1^{-1}, t_2^{-1}, \cdots, t_n^{-1}\right) B$ is a γ - cut

Remark: Computing γ - cut requires litter extra computation

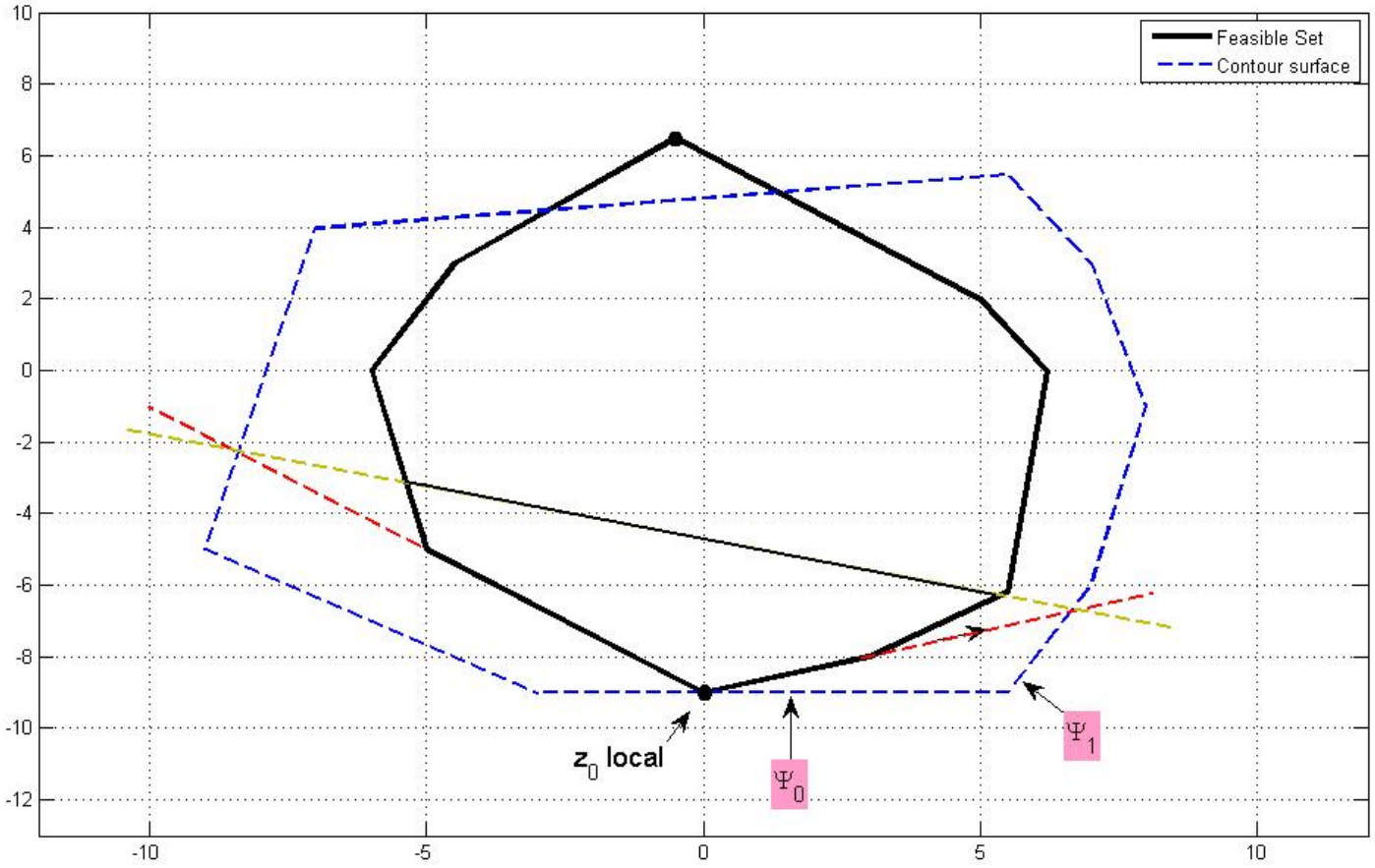
4.2. Utilizing γ - cut in hill-detouring

The hill-detouring process may be implemented by solving the following PL optimization problem

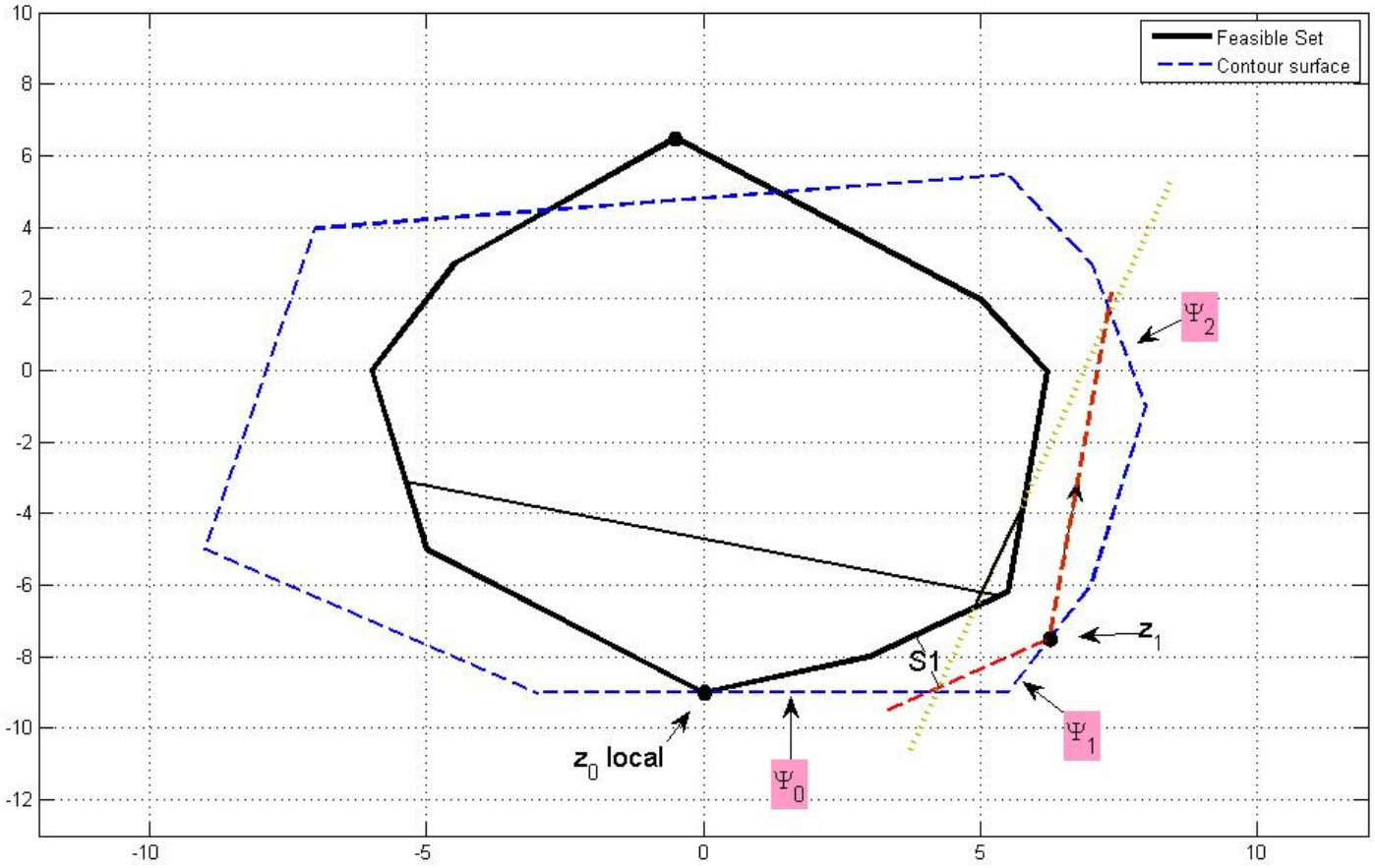
$$\begin{aligned} \min \quad & \lambda \\ \text{s.t.} \quad & a_i^T x - b_i \le \lambda, \ \ i = 1, 2, \cdots, M \\ & \alpha_0^T \left(x - z_0 \right) \ge 1 \\ & f\left(x \right) \le f\left(z_0 \right) \end{aligned} \qquad \textbf{(P0)}$$

Utlizing the first property mentioned above, a locally optimal solution $\left(\lambda_1, z_1 \right)$ can be found by means of simplex algorithm, starting from a γ - extension of z_0. If $\lambda_1 \le 0$, the hill-detouring process is completed. Otherwise, a γ - cut may be produced for later hill-detouring process. This process is illustrated below

Feasible set and contour surface revised by γ - cut at z_0



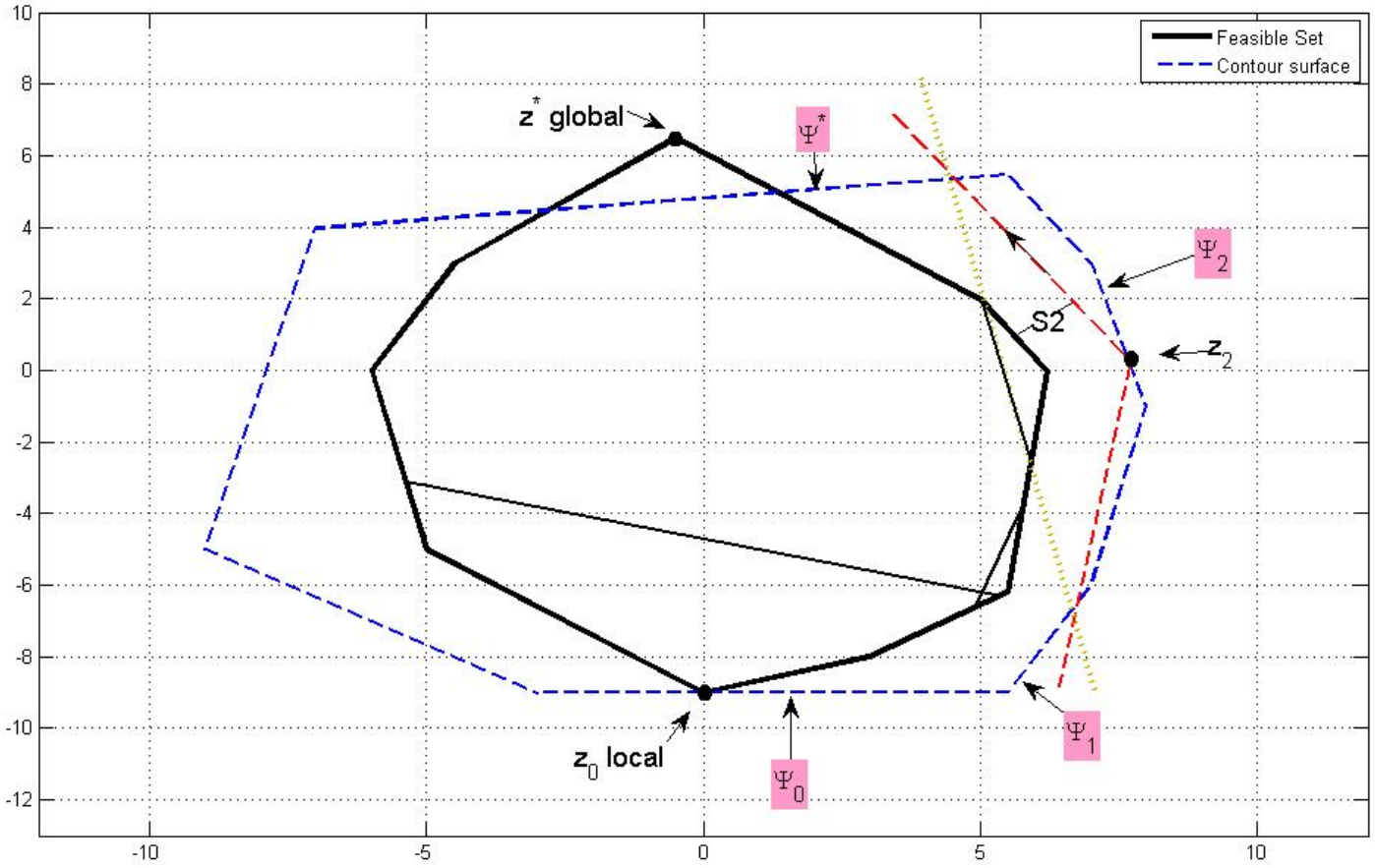
Locally optimal point z_1 of (P0) and corresponding γ - cut



Utilizing γ - cut at z_1, say $\alpha_1^T \left(x - z_1 \right) = 1$, the PL optimization problem for hill-detouring becomes

$$
\begin{aligned}
\min \quad & \lambda \\
\text{s.t.} \quad & a_i^T x - b_i \le \lambda, \quad i = 1, 2, \cdots, M \\
& \alpha_0^T \left(x - z_0 \right) \ge 1 \\
& \alpha_1^T \left(x - z_1 \right) \ge 1 \\
& f\left(x \right) \le f\left(z_0 \right)
\end{aligned}
\qquad \textbf{(P1)}
$$

Locally optimal point z_2 of (P1) and corresponding γ - cut



Utilizing γ - cut at z_2, say $\alpha_2^T\left(x-z_2\right)=1$, the PL optimization problem for hill-detouring becomes

$$
\begin{aligned}
&\min \quad \lambda \\
&\text{s.t.} \quad a_i^T x - b_i \le \lambda, \quad i = 1,2,\cdots,M \\
&\qquad \alpha_0^T\left(x-z_0\right) \ge 1 \\
&\qquad \alpha_1^T\left(x-z_1\right) \ge 1 \\
&\qquad \alpha_2^T\left(x-z_2\right) \ge 1 \\
&\qquad f\left(x\right) \le f\left(z_0\right)
\end{aligned}
\qquad \textbf{(P2)}
$$

The hill-detouring process is terminated after solving (P2)

Two favorable facts

i. Active feasible region and contour surface are continuously reduced, which excludes all known locally optimal points, hence cycle will not happen

ii. Once the remaining feasible region becomes empty, which may be checked during the optimization process, z_0 has been a globally optimal solution of the original cancave PL optimization problem

Convergence of the algorithm may be ensured by these facts.

Numerical comparison

Testing problem: minimizing hinging hyperplanes on hypercube

$$\min \quad c_0^T x + \sum_{i=1}^{M} \sigma_i \max\left\{0, c_i^T x + d_i\right\}$$

$$\text{s.t.} \quad 0 \le x_j \le 1, \quad j = 1, 2, \cdots, n$$

where $\sigma_i \in \{-1, 1\}, \ i = 1, 2, \cdots, M$

Corresponding concave problem

$$\min \quad c_0^T x + \sum_{\sigma_i = 1} y_i - \sum_{\sigma_i = -1} \max\left\{0, c_i^T x + d_i\right\}$$

$$\text{s.t.} \quad c_i^T x + d_i \le y_i, \quad y_i \ge 0, \quad \forall \sigma_i = 1$$

$$0 \le x_j \le 1, \quad j = 1, 2, \cdots, n$$

Numerical results

Table 2: Computations for numerical examples via three algorithms

M=100	MIP		HD			GHD			Reduction
Dim	Value	Time	Value	Time	T_e	Value	Time	T_e	%
3	-2.423	23.750	-2.422	12.558	2	-2.422	1.952	6	84.5
3	-7.535	2.625	-7.535	8.016	0	-7.535	0.129	0	98.4
3	-5.274	1.453	-5.274	11.084	0	-5.274	0.105	0	99.0
3	-4.028	11.500	-4.029	8.577	0	-4.029	0.612	2	92.9
3	-3.764	4.953	-3.765	8.450	0	-3.764	1.044	3	87.6
3	-6.413	3.110	-6.413	8.000	0	-6.413	0.439	1	94.5
3	-4.052	18.875	-4.052	19.636	0	-4.052	0.072	0	99.6
3	-5.873	9.422	-5.783	18.512	3	-5.873	0.336	1	98.2
3	-5.370	6.079	-5.370	8.884	1	-5.370	0.872	2	90.2
5	-13.271	230.156	-13.27	10.820	0	-13.271	0.072	0	99.3
5	-8.217	364.360	-8.216	24.447	0	-8.217	0.354	1	98.6
5	-3.989	1000.000	-3.989	10.606	0	-3.989	0.498	1	95.3
5	-7.049	185.844	-6.993	42.652	0	-7.049	0.624	2	98.5
5	-8.288	229.672	-8.288	23.989	1	-8.287	0.100	0	99.6
5	-7.870	452.140	-7.870	20.416	3	-7.870	0.092	0	99.5
5	-16.189	279.360	-16.189	9.824	0	-16.189	0.063	0	99.4
5	-13.601	40.031	-13.60	21.690	1	-13.601	0.348	1	98.4
5	-10.422	1000.000	-10.42	16.961	1	-10.422	0.315	1	98.1
10	-15.102	1000.000	-15.10	41.961	1	-15.103	1.269	3	96.9
10	-15.105	1000.000	-15.10	35.572	0	-15.105	0.625	1	95.9
10	-18.840	1001.031	-18.76	69.999	3	-18.245	2.399	9	99.6
10	-11.436	1000.016	-11.44	35.580	2	-11.436	0.633	1	98.2
10	-24.519	1000.531	-24.519	44.675	2	-24.378	0.123	0	99.7
10	-6.397	1000.422	-6.397	36.028	1	-6.396	0.548	1	98.5
10	-8.590	1000.031	-8.590	43.291	6	-8.590	1.056	2	97.6
10	-9.187	1000.000	-9.187	32.272	2	-9.187	0.897	1	97.2

5. Conclusions

i. Some classical DDM problems can be transformed to concave PL optimization problem

ii. Hill-detouring is a practical way to solve concave PL optimization problem

iii. Utilizing γ - cut may remarkably improve the performance of hill-detouring approach

Thanks!

Questions?