The Cluster Problem in Constrained Global Optimization

Rohit Kannan and Paul I. Barton

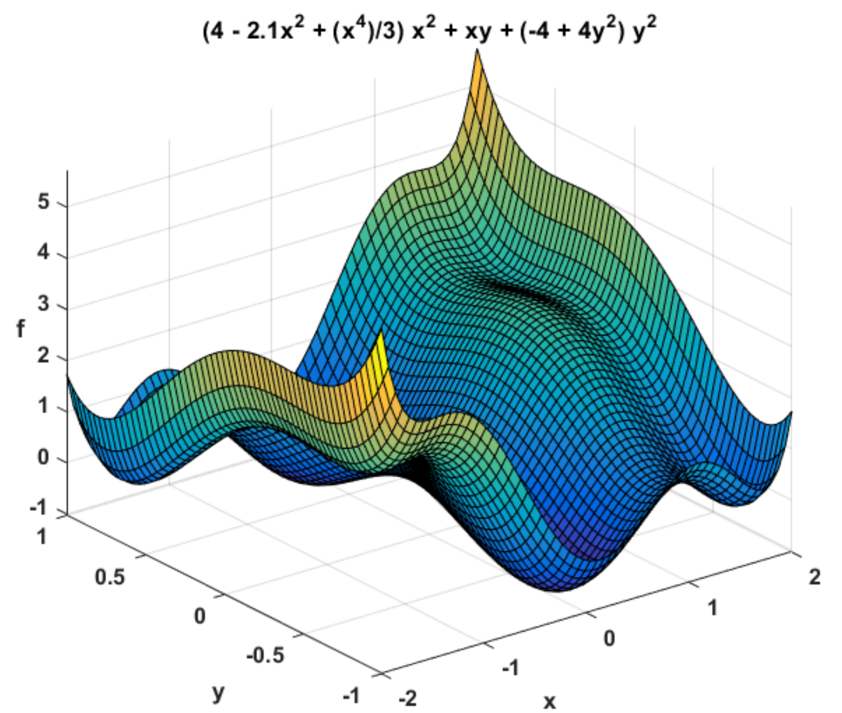
Process Systems Engineering Laboratory
Department of Chemical Engineering
Massachusetts Institute of Technology

September 08, 2016

Motivation

Clustering in Unconstrained Optimization

- Consider the unconstrained minimization of

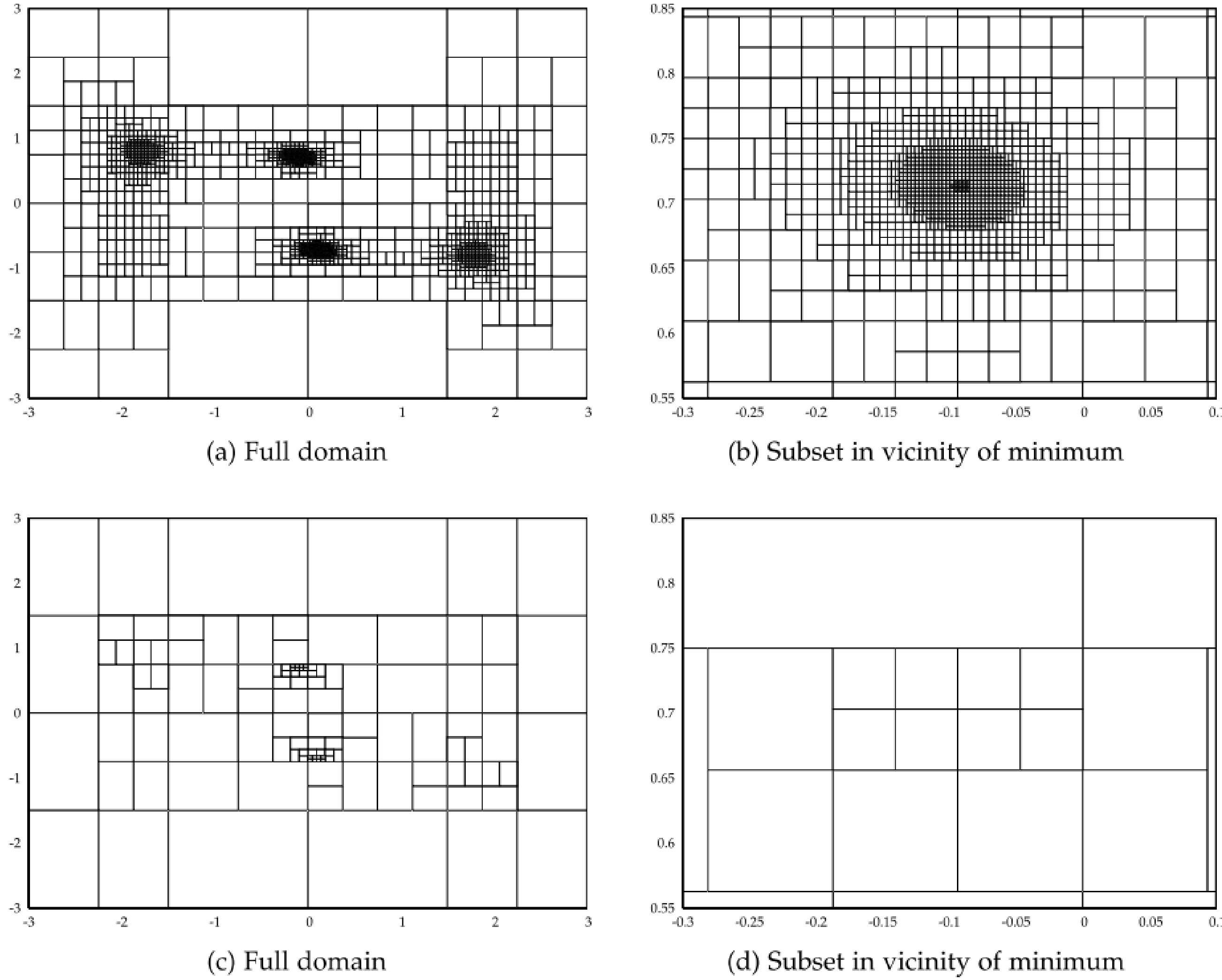


Motivation

Clustering in Unconstrained Optimization

Natural Interval Extension

Centered Form



(a) Full domain

(b) Subset in vicinity of minimum

(c) Full domain

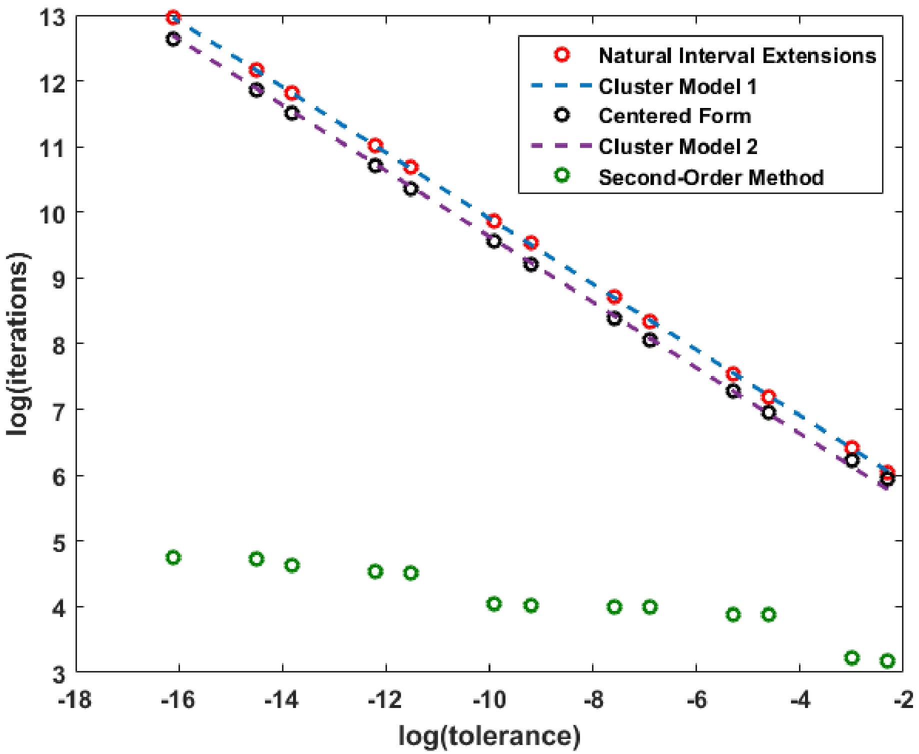
(d) Subset in vicinity of minimum

Wechsung, A., Ph.D. Thesis, MIT, 2014.

Motivation

Clustering in Constrained Optimization

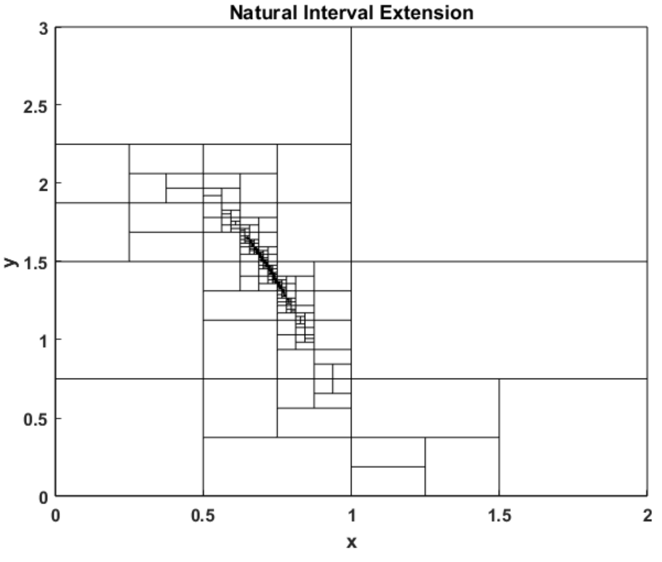
$$\min_{x,y}\ y^2 - 12x - 7y$$
$$\text{s.t. } y + 2x^4 - 2 = 0,$$
$$x \in [0,2],\ y \in [0,3].$$

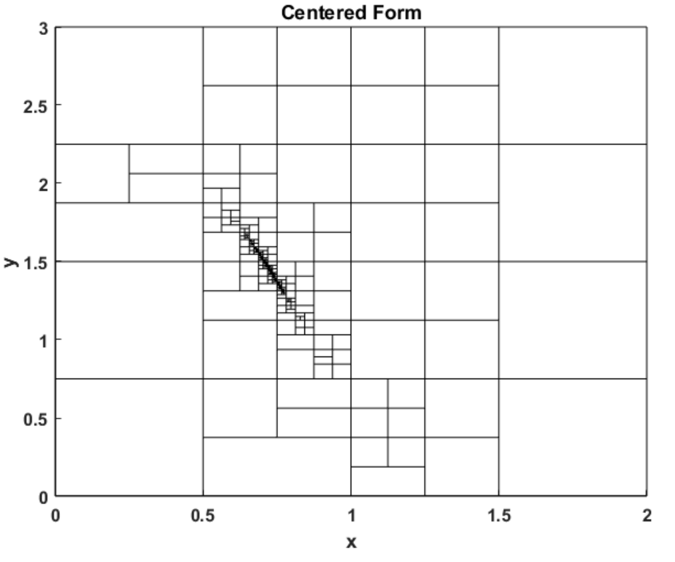


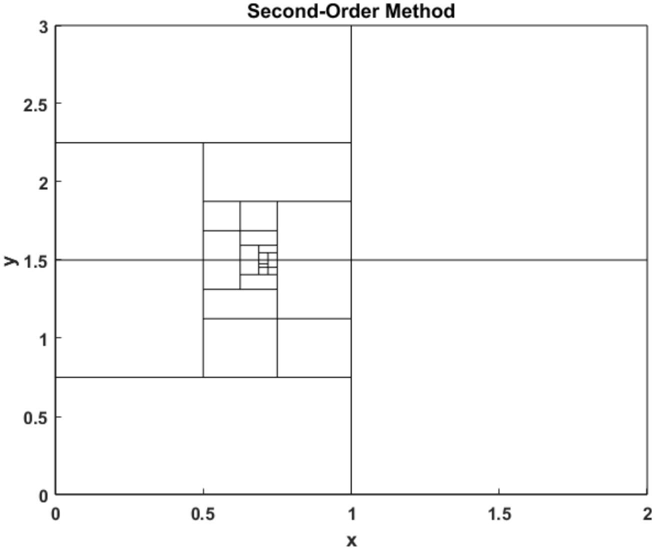
Floudas, C. et al., Springer, 1999.

Motivation

Clustering in Constrained Optimization

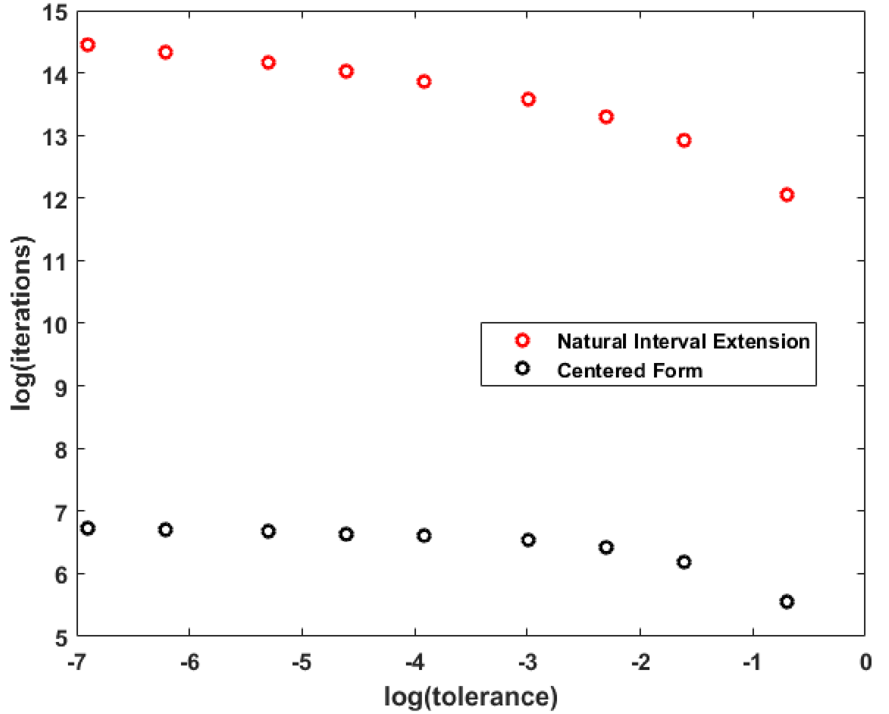






Motivation
Clustering in Constrained Optimization

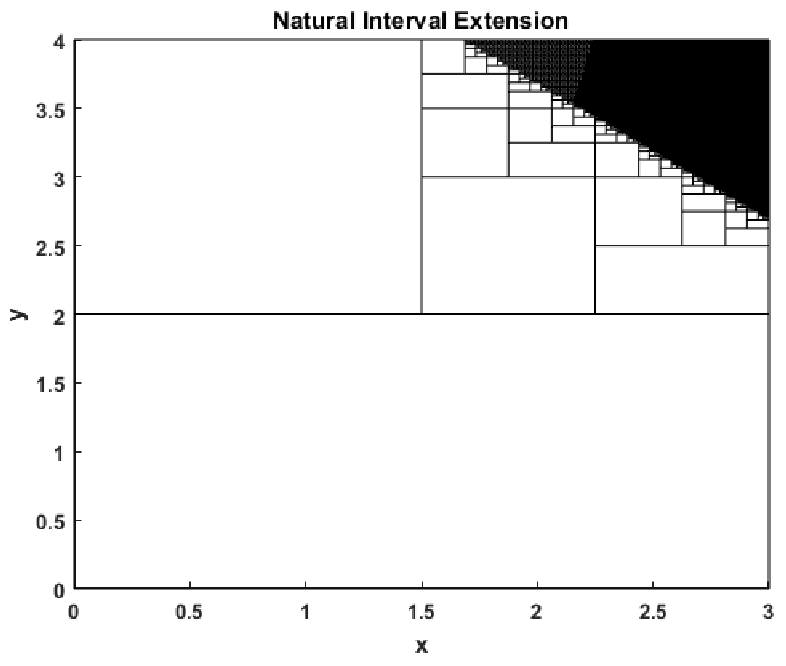
$$\min_{x,y} \; -x-y$$
$$\text{s.t.} \;\; y \leq 2+2x^4-8x^3+8x^2,$$
$$y \leq 4x^4-32x^3+88x^2-96x+36,$$
$$x \in [0,3], \; y \in [0,4].$$

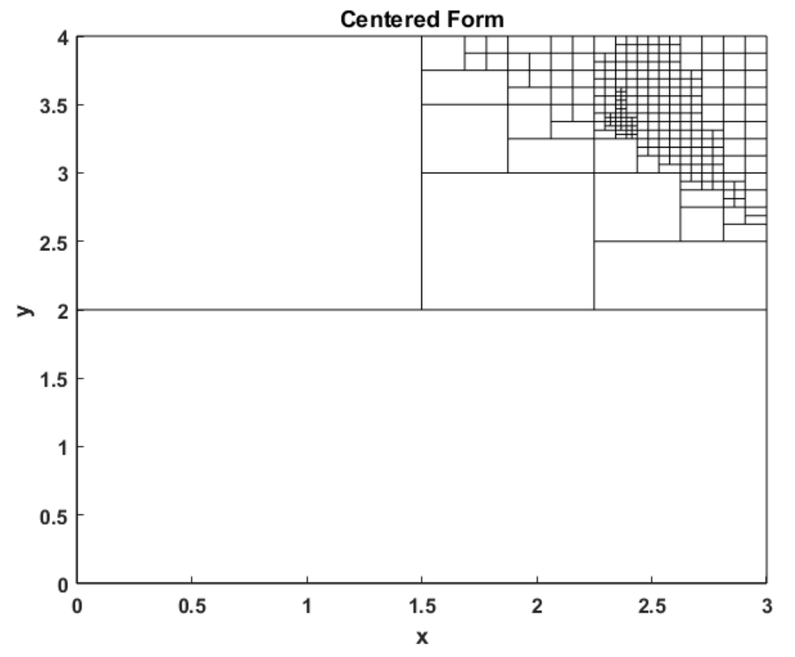


Floudas, C. et al., Springer, 1999.

Motivation

Clustering in Constrained Optimization



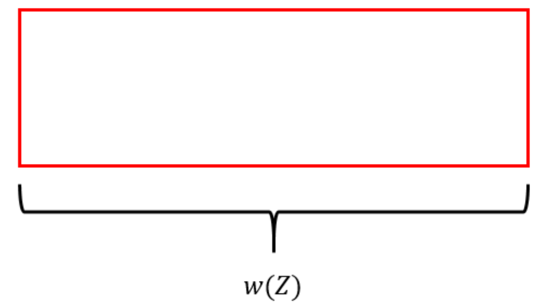


Definitions

- Width of an interval

Let $Z = [z_1^{\mathrm{L}}, z_1^{\mathrm{U}}] \times \cdots \times [z_n^{\mathrm{L}}, z_n^{\mathrm{U}}] \in \mathbb{IR}^n$.

The width of Z is given by $w(Z) = \max\limits_{i=1,\cdots,n} (z_i^{\mathrm{U}} - z_i^{\mathrm{L}})$.



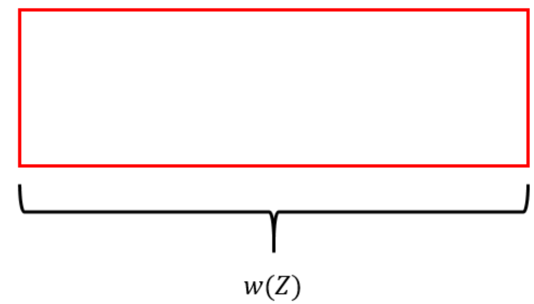
Bompadre, A. et al., J. Global Optim., 2012.

Definitions

- **Width of an interval**

Let $Z = [z_1^{\mathrm{L}}, z_1^{\mathrm{U}}] \times \cdots \times [z_n^{\mathrm{L}}, z_n^{\mathrm{U}}] \in \mathbb{IR}^n$.

The width of Z is given by $w(Z) = \max\limits_{i=1,\cdots,n}(z_i^{\mathrm{U}} - z_i^{\mathrm{L}})$.



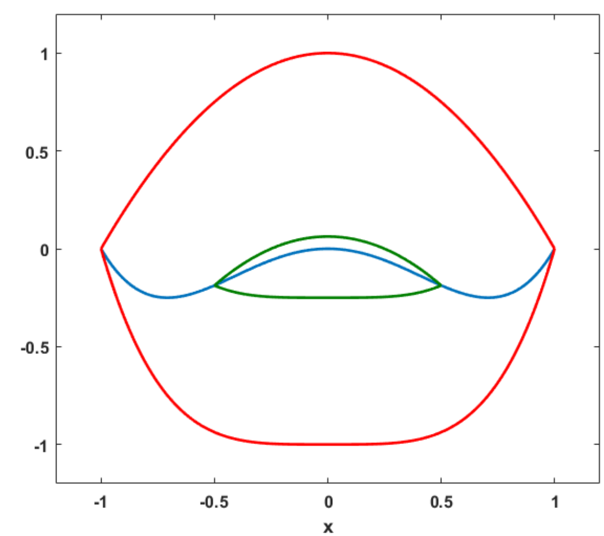
- **Schemes of relaxations**

Nonempty, bounded set $X \subset \mathbb{R}^n$, function $h : X \to \mathbb{R}$.

For each interval $Z \in \mathbb{I}X$, define convex relaxation $h_Z^{\mathrm{cv}} : Z \to \mathbb{R}$, concave relaxation $h_Z^{\mathrm{cc}} : Z \to \mathbb{R}$.

$\left(h_Z^{\mathrm{cv}}\right)\Big|_{Z \in \mathbb{I}X}$ defines a scheme of convex relaxations of h in X.

$\left(h_Z^{\mathrm{cc}}\right)\Big|_{Z \in \mathbb{I}X}$ defines a scheme of concave relaxations of h in X.



Bompadre, A. et al., J. Global Optim., 2012.

Definitions

- Hausdorff metric

Suppose $X = [x^{\mathrm{L}}, x^{\mathrm{U}}], Y = [y^{\mathrm{L}}, y^{\mathrm{U}}] \in \mathbb{IR}$ are two intervals.

Hausdorff metric $q(X, Y) := \max\left\{\left|x^{\mathrm{L}} - y^{\mathrm{L}}\right|, \left|x^{\mathrm{U}} - y^{\mathrm{U}}\right|\right\}$.

Definitions

- ## Hausdorff metric

Suppose $X = [x^{\mathrm{L}}, x^{\mathrm{U}}], Y = [y^{\mathrm{L}}, y^{\mathrm{U}}] \in \mathbb{IR}$ are two intervals.

Hausdorff metric $q(X, Y) := \max\left\{\left|x^{\mathrm{L}} - y^{\mathrm{L}}\right|, \left|x^{\mathrm{U}} - y^{\mathrm{U}}\right|\right\}$.

- ## Inclusion function

$h : \mathbb{R}^n \supset X \to \mathbb{R}$ continuous.

Image of $Z \subset X$ under h: $\bar{h}(Z) := \left[h^{\mathrm{L}}(Z), h^{\mathrm{U}}(Z)\right]$.

$H : \mathbb{I}X \supset \mathcal{X} \to \mathbb{IR}$ is an inclusion function for h on $\mathcal{X}$ if

$$\bar{h}(Z) \subset H(Z), \forall Z \in \mathcal{X}.$$

Definitions

- **Hausdorff metric**

Suppose $X = [x^{\mathrm{L}}, x^{\mathrm{U}}], Y = [y^{\mathrm{L}}, y^{\mathrm{U}}] \in \mathbb{IR}$ are two intervals.

Hausdorff metric $q(X, Y) := \max\left\{\left|x^{\mathrm{L}} - y^{\mathrm{L}}\right|, \left|x^{\mathrm{U}} - y^{\mathrm{U}}\right|\right\}$.

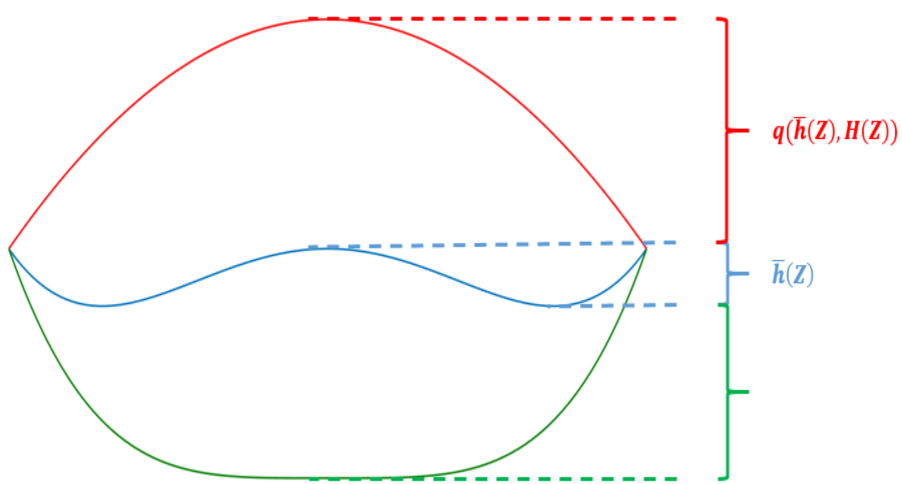
- **Inclusion function**

$h : \mathbb{R}^n \supset X \to \mathbb{R}$ continuous.

Image of $Z \subset X$ under h: $\bar{h}(Z) := \left[h^{\mathrm{L}}(Z), h^{\mathrm{U}}(Z)\right]$.

$H : \mathbb{I}X \supset \mathcal{X} \to \mathbb{IR}$ is an inclusion function for h on $\mathcal{X}$ if

$$\bar{h}(Z) \subset H(Z), \forall Z \in \mathcal{X}.$$



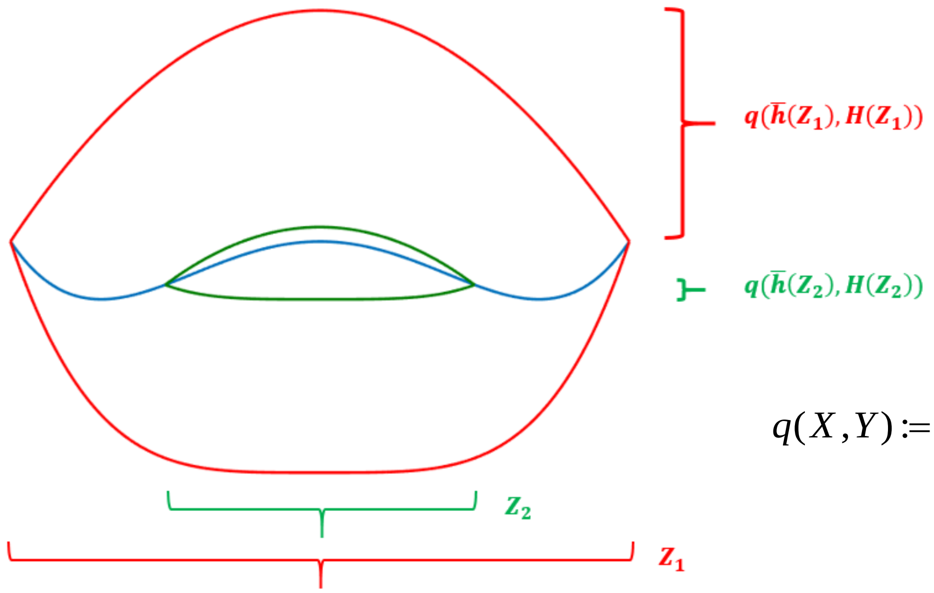
Hausdorff Convergence

◆ Hausdorff Convergence Order

$h:\mathbb{R}^n \supset X \to \mathbb{R}$ continuous, H inclusion function of h on $\mathbb{I}X$.

H has Hausdorff convergence of order $\beta > 0$ on X if $\exists \tau > 0$ s.t. $\forall Z \in \mathbb{I}X$,

$$q(\bar{h}(Z), H(Z)) \leq \tau w(Z)^{\beta}.$$



$$q(X,Y) := \max\left\{\left|x^{\mathrm{L}} - y^{\mathrm{L}}\right|, \left|x^{\mathrm{U}} - y^{\mathrm{U}}\right|\right\}.$$

Bompadre, A. et al., J. Global Optim., 2012.

Pointwise Convergence

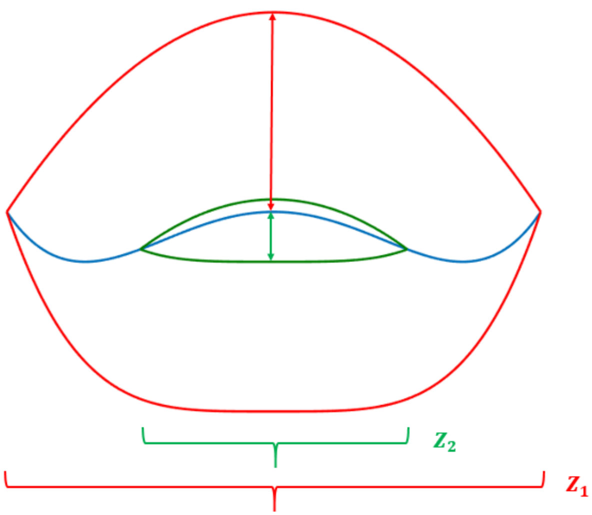
◆ Pointwise Convergence Order

$h:\mathbb{R}^n \supset X \to \mathbb{R}$ continuous, $(h_Z^{cv}, h_Z^{cc})\big|_{Z\in\mathbb{I}X}$ scheme of relaxations of h in X.

$(h_Z^{cv}, h_Z^{cc})\big|_{Z\in\mathbb{I}X}$ has pointwise convergence of order $\gamma > 0$ on X if $\exists\,\tau > 0$ s.t. $\forall Z \in \mathbb{I}X$,

$$\sup_{x\in Z}\left|h(x) - h_Z^{cv}(x)\right| \le \tau w(Z)^{\gamma},$$

$$\sup_{x\in Z}\left|h(x) - h_Z^{cc}(x)\right| \le \tau w(Z)^{\gamma}.$$



Bompadre, A. et al., J. Global Optim., 2012.

Propagation of convergence orders

- γ-order pointwise convergence of a scheme of relaxations implies $(\gamma \leq)\beta$-order Hausdorff convergence of the scheme
- Envelopes and αBB relaxations have second-order pointwise convergence for C^2 functions
- Natural interval extensions have first-order pointwise convergence for Lipschitz continuous functions
- Centered forms have second-order Hausdorff convergence for C^1 functions

Bompadre, A. et al., J. Global Optim., 2012.

Scholz, D., J. Global Optim., 2012.

Propagation of convergence orders

Convergence order of factors	Convergence order of operation result
Sum: $g(\mathbf{z}) = g_1(\mathbf{z}) + g_2(\mathbf{z})$	
Schemes for g_i have β_i	$\beta \geq 1$ (no order propagation)
Schemes for g_i have γ_i	$\gamma \geq \min\{\gamma_1, \gamma_2\}$
Product: $g(\mathbf{z}) = g_1(\mathbf{z}) \cdot g_2(\mathbf{z})$	
Schemes for g_i have β_i	$\beta \geq 1$ (no order propagation)
Schemes for g_i have γ_i	$\gamma \geq \min\{\gamma_1, \gamma_2, 2\}$
Composition: $g(\mathbf{z}) = F \circ f(\mathbf{z})$	
Scheme for F has β_F	$\beta \geq \min\{\beta_F, \beta_{f,T}\}$
Inclusion for f has $\beta_{f,T}$	
Scheme for F has γ_F	$\gamma \geq \min\{\gamma_F, \gamma_f\}$
Scheme for f has γ_f	

Bound on convergence order of McCormick estimators assuming Lipschitz continuity of the factors

Bompadre, A. et al., J. Global Optim., 2012.

Bompadre, A. et al., J. Global Optim., 2013.

More Definitions

- Distance between sets

Let $Y, Z \subset \mathbb{R}^n$.

The distance between Y and Z is defined as

$$d(Y, Z) := \inf_{\substack{y \in Y, \\ z \in Z}} \|y - z\|.$$

More Definitions

- Distance between sets

Let $Y, Z \subset \mathbb{R}^n$.

The distance between Y and Z is defined as

$$d(Y,Z) := \inf_{\substack{y\in Y,\\ z\in Z}} \|y - z\|.$$

- Convergence and Pointwise Convergence

$h : \mathbb{R}^n \supset X \to \mathbb{R}$ continuous, $(h_Z^{\text{cv}})\big|_{Z\in\mathbb{I}X}$ scheme of convex relaxations of h on X.

$(h_Z^{\text{cv}})\big|_{Z\in\mathbb{I}X}$ has convergence of order $\beta > 0$ on X if $\exists \tau > 0$ s.t. $\forall Z \in \mathbb{I}X$,

$$\inf_{x\in Z} h(x) - \inf_{x\in Z} h_Z^{\text{cv}}(x) \le \tau w(Z)^\beta.$$

$(h_Z^{\text{cv}})\big|_{Z\in\mathbb{I}X}$ has pointwise convergence of order $\gamma > 0$ on X if $\exists \tau > 0$ s.t. $\forall Z \in \mathbb{I}X$,

$$\sup_{x\in Z} \left| h(x) - h_Z^{\text{cv}}(x) \right| \le \tau w(Z)^\gamma.$$

Clustering in Unconstrained Global Optimization

Suppose

- $X \subset \mathbb{R}^n$ is an open, convex set
- $f : X \to \mathbb{R}$ is $\mathcal{C}^2$ on X

Du, K. and Kearfott, R.B., J. Global Optim., 1994.

Neumaier, A., Acta Numerica, 2004.

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Suppose

- $X \subset \mathbb{R}^n$ is an open, convex set
- $f : X \to \mathbb{R}$ is $\mathcal{C}^2$ on X
- x^* is the unique unconstrained global minimum of f on X
- $\nabla^2 f(x^*)$ is positive definite

Du, K. and Kearfott, R.B., J. Global Optim., 1994.

Neumaier, A., Acta Numerica, 2004.

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Suppose

- $X \subset \mathbb{R}^n$ is an open, convex set
- $f : X \to \mathbb{R}$ is $\mathcal{C}^2$ on X
- x^* is the unique unconstrained global minimum of f on X
- $\nabla^2 f(x^*)$ is positive definite
- The B&B algorithm finds the upper bound $UBD = f(x^*)$ early on
- The termination tolerance $\varepsilon \ll 1$
- The B&B algorithm terminates when $UBD - LBD \leq \varepsilon$

Du, K. and Kearfott, R.B., J. Global Optim., 1994.

Neumaier, A., Acta Numerica, 2004.

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Suppose

- $X \subset \mathbb{R}^n$ is an open, convex set
- $f : X \to \mathbb{R}$ is $\mathcal{C}^2$ on X
- x^* is the unique unconstrained global minimum of f on X
- $\nabla^2 f(x^*)$ is positive definite
- The B&B algorithm finds the upper bound $UBD = f(x^*)$ early on
- The termination tolerance $\varepsilon \ll 1$
- The B&B algorithm terminates when $UBD - LBD \leq \varepsilon$
- The scheme of convex relaxations $(f_Z^{\text{cv}})\big|_{Z \in \mathbb{I}X}$ has convergence of order $\beta > 0$ on X

Du, K. and Kearfott, R.B., J. Global Optim., 1994.

Neumaier, A., Acta Numerica, 2004.

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Let $\delta = \left(\frac{\varepsilon}{\tau} \right)^{\frac{1}{\beta}}$.

Partition X into regions A and B such that

$$A = \left\{ x \in X : f(x) - f(x^*) > \varepsilon \right\},$$

$$B = \left\{ x \in X : f(x) - f(x^*) \leq \varepsilon \right\}.$$

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Let $\delta = \left(\frac{\varepsilon}{\tau}\right)^{\frac{1}{\beta}}$.

Partition X into regions A and B such that

$$A = \left\{ x \in X : f(x) - f(x^*) > \varepsilon \right\},$$

$$B = \left\{ x \in X : f(x) - f(x^*) \le \varepsilon \right\}.$$

If $Z \in \mathbb{I}A$,

$$\min_{x \in Z} f(x) - \min_{x \in Z} f_Z^{\mathrm{cv}}(x) \le \tau w(Z)^{\beta}$$

$$\Rightarrow \min_{x \in Z} f_Z^{\mathrm{cv}}(x) \ge \min_{x \in Z} f(x) - \tau w(Z)^{\beta} > f(x^*) + \varepsilon - \tau w(Z)^{\beta}$$

$$\therefore \min_{x \in Z} f_Z^{\mathrm{cv}}(x) \ge f(x^*) - \varepsilon \text{ when } \tau w(Z)^{\beta} \le 2\varepsilon \Leftrightarrow w(Z) \le 2^{\frac{1}{\beta}} \delta$$

Condition for fathoming

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Let $\delta = \left(\frac{\varepsilon}{\tau}\right)^{\frac{1}{\beta}}$.

Partition X into regions A and B such that

$$A = \{x \in X : f(x) - f(x^*) > \varepsilon\},$$

$$B = \{x \in X : f(x) - f(x^*) \le \varepsilon\}.$$

If $Z \in \mathbb{I}A$,

$$\min_{x \in Z} f(x) - \min_{x \in Z} f_Z^{\text{cv}}(x) \le \tau w(Z)^\beta$$

$$\Rightarrow \min_{x \in Z} f_Z^{\text{cv}}(x) \ge \min_{x \in Z} f(x) - \tau w(Z)^\beta > f(x^*) + \varepsilon - \tau w(Z)^\beta$$

$$\therefore \min_{x \in Z} f_Z^{\text{cv}}(x) \ge f(x^*) - \varepsilon \text{ when } \tau w(Z)^\beta \le 2\varepsilon \Leftrightarrow w(Z) \le 2^{\frac{1}{\beta}} \delta$$

If $Z \in \mathbb{I}B$,

$$\min_{x \in Z} f(x) - \min_{x \in Z} f_Z^{\text{cv}}(x) \le \tau w(Z)^\beta$$

$$\Rightarrow \min_{x \in Z} f_Z^{\text{cv}}(x) \ge \min_{x \in Z} f(x) - \tau w(Z)^\beta \ge f(x^*) - \tau w(Z)^\beta$$

$$\therefore \min_{x \in Z} f_Z^{\text{cv}}(x) \ge f(x^*) - \varepsilon \text{ when } \tau w(Z)^\beta \le \varepsilon \Leftrightarrow w(Z) \le \delta$$

Condition for fathoming

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

Let $\delta = \left(\dfrac{\varepsilon}{\tau}\right)^{\frac{1}{\beta}}$.

Partition X into regions A and B such that

$$A = \left\{ x \in X : f(x) - f(x^*) > \varepsilon \right\},$$

$$B = \left\{ x \in X : f(x) - f(x^*) \leq \varepsilon \right\}.$$

$$B = \left\{ x \in X : f(x) - f(x^*) \leq \varepsilon \right\}$$

$$\approx \left\{ x \in X : \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(x^*)(x - x^*) \leq \varepsilon \right\}$$

Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

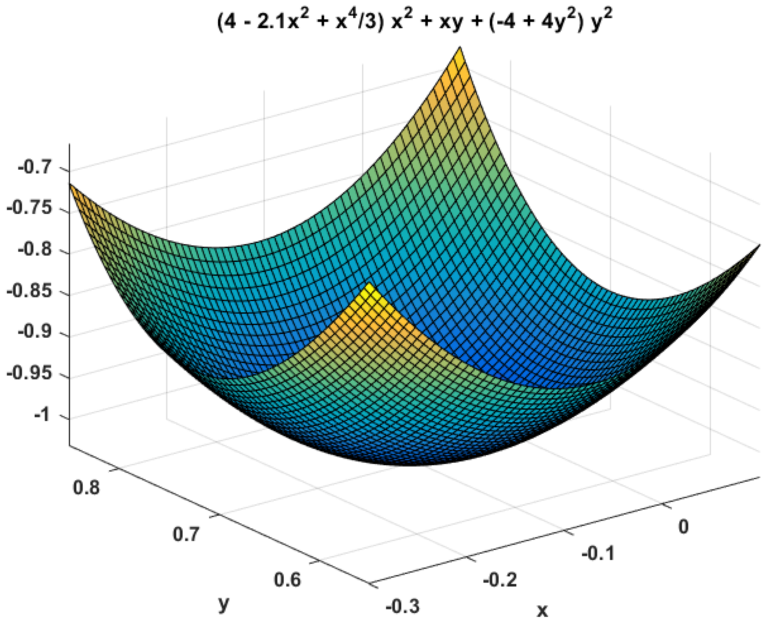
Let $\delta = \left(\frac{\varepsilon}{\tau}\right)^{\frac{1}{\beta}}$.

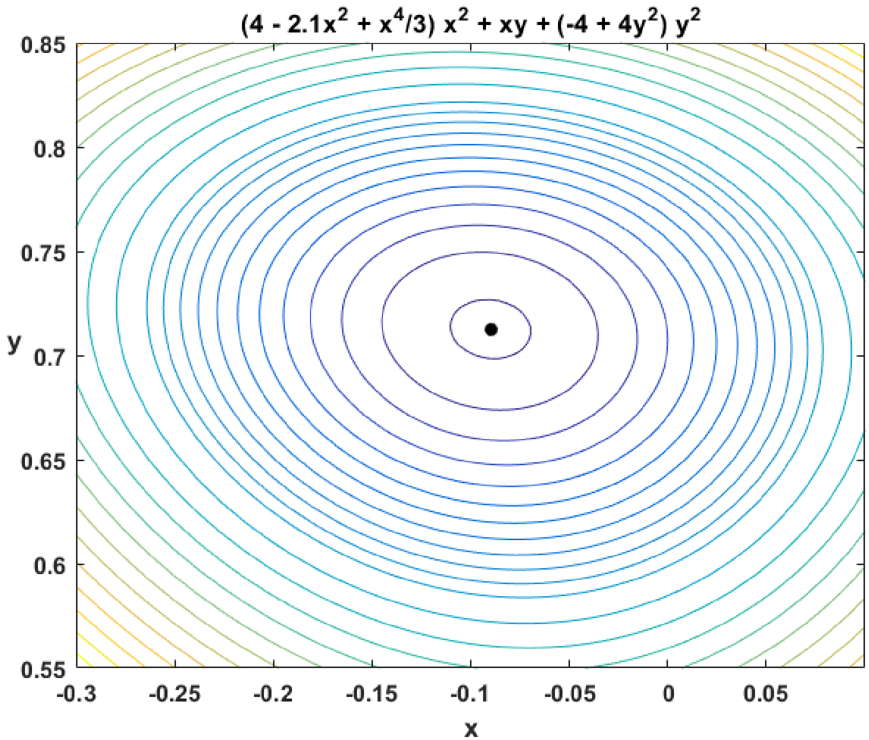
Partition X into regions A and B such that

$$A = \left\{x \in X : f(x) - f(x^*) > \varepsilon\right\},$$

$$B = \left\{x \in X : f(x) - f(x^*) \leq \varepsilon\right\}.$$

$$B = \left\{x \in X : f(x) - f(x^*) \leq \varepsilon\right\}$$
$$\approx \left\{x \in X : \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(x^*)(x - x^*) \leq \varepsilon\right\}$$





Wechsung, A. et al., J. Global Optim., 2014.

Clustering in Unconstrained Global Optimization

$\delta = \left(\frac{\varepsilon}{\tau}\right)^{\frac{1}{\beta}}$. $B \approx \left\{ x \in X : \frac{1}{2}(x - x^*)^{\mathrm{T}} \nabla^2 f(x^*)(x - x^*) \leq \varepsilon \right\}$. $\lambda_1 =$ smallest eigenvalue of $\nabla^2 f(x^*)$.

Cover B using boxes of width δ to estimate the extent of clustering.

$(f_Z^{\mathrm{cv}})\big|_{Z \in \mathbb{I}X}$ has convergence of order $\beta > 0$ on X, i.e., $\exists \tau > 0$ s.t. $\forall Z \in \mathbb{I}X$,

$$\min_{x \in Z} f(x) - \min_{x \in Z} f_Z^{\mathrm{cv}}(x) \leq \tau w(Z)^{\beta}.$$

Clustering in Unconstrained Global Optimization

$\delta = \left(\frac{\varepsilon}{\tau}\right)^{\frac{1}{\beta}}$. $B \approx \left\{ x \in X : \frac{1}{2}(x-x^*)^{\mathrm{T}} \nabla^2 f(x^*)(x-x^*) \le \varepsilon \right\}$. λ_1 = smallest eigenvalue of $\nabla^2 f(x^*)$.

Cover B using boxes of width δ to estimate the extent of clustering.

$(f_Z^{\mathrm{cv}})\big|_{Z \in \mathbb{I}X}$ has convergence of order $\beta > 0$ on X, i.e., $\exists \tau > 0$ s.t. $\forall Z \in \mathbb{I}X$,

$$\min_{x \in Z} f(x) - \min_{x \in Z} f_Z^{\mathrm{cv}}(x) \le \tau w(Z)^{\beta}.$$

Number of boxes required to cover B when $\beta = 2$ (Wechsung et al., 2014)

Case	Number of boxes
$\tau \le \frac{\lambda_1}{8}$	1
$\frac{\lambda_1}{8} < \tau \le \frac{2\lambda_1}{8}$	$1 + 2n$
$\frac{2\lambda_1}{8} < \tau \le \frac{3\lambda_1}{8}$	$1 + 2n^2$
$\frac{3\lambda_1}{8} < \tau \le \frac{4\lambda_1}{8}$	$1 + \frac{8}{3}n - 2n^2 + \frac{4}{3}n^3$
$\vdots$	$\vdots$
$\frac{18\lambda_1}{8} < \tau$	$\left\lceil 2\sqrt{\tau\lambda_1^{-1}} \right\rceil^{n-1} \left(\left\lceil 2\sqrt{\tau\lambda_1^{-1}} \right\rceil + 2n \left\lceil (\sqrt{2}-1)\sqrt{\tau\lambda_1^{-1}} \right\rceil \right)$

Formulation

Consider the problem

$$\begin{aligned}\min_{x \in X}\ & f(x) \\ \text{s.t. } & g(x) \le 0, \\ & h(x) = 0,\end{aligned}$$

where $X \subset \mathbb{R}^n$ is a nonempty open bounded convex set, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^{m_I}$, $h : X \to \mathbb{R}^{m_E}$.

Formulation

Consider the problem

$$\min_{x \in X} f(x)$$
$$\text{s.t. } g(x) \le 0,$$
$$h(x) = 0,$$

where $X \subset \mathbb{R}^n$ is a nonempty open bounded convex set, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^{m_I}$, $h : X \to \mathbb{R}^{m_E}$.

Assume

1. $f, g,$ and h are $\mathcal{C}^2$ on X
2. The constraints define a compact set inside X

Formulation

Consider the problem

$$\begin{aligned}\min_{x\in X}\ & f(x)\\ \text{s.t. } & g(x)\le 0,\\ & h(x)=0,\end{aligned}$$

where $X \subset \mathbb{R}^n$ is a nonempty open bounded convex set, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^{m_I}$, $h : X \to \mathbb{R}^{m_E}$.

Assume

1. $f, g,$ and h are $\mathcal{C}^2$ on X
2. The constraints define a compact set inside X
3. $x^* \in X$ is a global minimum of the above problem, and the B&B algorithm has found the upper bound $UBD = f(x^*)$ early on
4. The termination tolerance $\varepsilon \ll 1$ and the algorithm fathoms node k when $UBD - LBD_k \le \varepsilon$

Convergence Order

Convex relaxation-based scheme

Let $(f_Z^{\mathrm{cv}})\big|_{Z\in\mathbb{I}X}$ and $(g_Z^{\mathrm{cv}})\big|_{Z\in\mathbb{I}X}$ denote continuous schemes of convex relaxations of f and g in X, and let $(h_Z^{\mathrm{cv}}, h_Z^{\mathrm{cc}})\big|_{Z\in\mathbb{I}X}$ denote a continuous scheme of relaxations of h in X.

Convergence Order
Convex relaxation-based scheme

Let $(f_Z^{\text{cv}})\big|_{Z\in\mathbb{I}X}$ and $(g_Z^{\text{cv}})\big|_{Z\in\mathbb{I}X}$ denote continuous schemes of convex relaxations of f and g in X, and let $(h_Z^{\text{cv}}, h_Z^{\text{cc}})\big|_{Z\in\mathbb{I}X}$ denote a continuous scheme of relaxations of h in X.

The convex relaxation-based lower bounding scheme is defined by

$$\begin{aligned}
\mathcal{O}(Z) &:= \min_{x\in Z} f_Z^{\text{cv}}(x)\\
&\qquad \text{s.t. } g_Z^{\text{cv}}(x)\le 0,\\
&\qquad\quad h_Z^{\text{cv}}(x)\le 0,\\
&\qquad\quad h_Z^{\text{cc}}(x)\ge 0,\\
\mathcal{I}_I(Z) &:= \overline{g}_Z^{\text{cv}}(Z),\\
\mathcal{I}_E(Z) &:= \left\{w\in\mathbb{R}^{m_E} : h_Z^{cv}(z)\le w\le h_Z^{cc}(z) \text{ for some } z\in Z\right\}.
\end{aligned}$$

$(\mathcal{O}(Z))\big|_{Z\in\mathbb{I}X}$: scheme of lower bounds.

$(\mathcal{I}_I(Z))\big|_{Z\in\mathbb{I}X}$: scheme estimating feasibility of inequality constraints.

$(\mathcal{I}_E(Z))\big|_{Z\in\mathbb{I}X}$: scheme estimating feasibility of equality constraints.

Convergence Order
Convex relaxation-based scheme

Let $\mathcal{F}(Z) := \{x \in Z : g(x) \leq 0, h(x) = 0\}$,

$$\mathcal{F}^{\mathrm{cv}}(Z) := \{x \in Z : g_Z^{\mathrm{cv}}(x) \leq 0, h_Z^{\mathrm{cv}}(x) \leq 0, h_Z^{\mathrm{cc}}(x) \geq 0\}.$$

The convex relaxation-based lower bounding scheme is said to have convergence of order $\beta > 0$ at

1. a feasible point $x \in X$ if $\exists \tau \geq 0$ s.t. $\forall Z \in \mathbb{I}X$ with $x \in Z$,

$$\min_{z \in \mathcal{F}(Z)} f(z) - \min_{z \in \mathcal{F}^{\mathrm{cv}}(Z)} f_Z^{\mathrm{cv}}(z) \leq \tau w(Z)^{\beta}.$$

2. an infeasible point $x \in X$ if $\exists \overline{\tau} \geq 0$ s.t. $\forall Z \in \mathbb{I}X$ with $x \in Z$,

$$d(\overline{g}(Z), \mathbb{R}_{-}^{m_I}) - d(\overline{g}_Z^{\mathrm{cv}}(Z), \mathbb{R}_{-}^{m_I}) \leq \overline{\tau} w(Z)^{\beta}, \text{ and}$$

$$d(\overline{h}(Z), \{0\}) - d(I_E(Z), \{0\}) \leq \overline{\tau} w(Z)^{\beta},$$

where $(I_E(Z))\big|_{Z \in \mathbb{I}X}$ is defined as

$$(I_E(Z))\big|_{Z \in \mathbb{I}X} := \left(\{w \in \mathbb{R}^{m_E} : h_Z^{\mathrm{cv}}(x) \leq w \leq h_Z^{\mathrm{cc}}(x) \text{ for some } x \in Z\}\right)_{Z \in \mathbb{I}X}.$$

Convergence Order

Convex relaxation-based scheme

Let $\mathcal{F}(Z) := \{x \in Z : g(x) \le 0, h(x) = 0\}$,

$\mathcal{F}^{\text{cv}}(Z) := \{x \in Z : g_Z^{\text{cv}}(x) \le 0, h_Z^{\text{cv}}(x) \le 0, h_Z^{\text{cc}}(x) \ge 0\}$.

The convex relaxation-based lower bounding scheme is said to have convergence of order $\beta > 0$ at

1. a feasible point $x \in X$ if $\exists \tau \ge 0$ s.t. $\forall Z \in \mathbb{I}X$ with $x \in Z$,

$$\min_{z \in \mathcal{F}(Z)} f(z) - \min_{z \in \mathcal{F}^{\text{cv}}(Z)} f_Z^{\text{cv}}(z) \le \tau w(Z)^{\beta}.$$

"The lower bound has to converge to the minimum objective value with order at least β"

2. an infeasible point $x \in X$ if $\exists \bar{\tau} \ge 0$ s.t. $\forall Z \in \mathbb{I}X$ with $x \in Z$,

$$d(\bar{g}(Z), \mathbb{R}_-^{m_I}) - d(\bar{g}_Z^{\text{cv}}(Z), \mathbb{R}_-^{m_I}) \le \bar{\tau} w(Z)^{\beta}, \text{ and}$$

$$d(\bar{h}(Z), \{0\}) - d(I_E(Z), \{0\}) \le \bar{\tau} w(Z)^{\beta},$$

where $(I_E(Z))\big|_{Z \in \mathbb{I}X}$ is defined as

$$(I_E(Z))\big|_{Z \in \mathbb{I}X} := \left(\left\{w \in \mathbb{R}^{m_E} : h_Z^{\text{cv}}(x) \le w \le h_Z^{\text{cc}}(x) \text{ for some } x \in Z\right\}\right)_{Z \in \mathbb{I}X}.$$

Convergence Order
Convex relaxation-based scheme

Let $\mathcal{F}(Z) := \left\{x \in Z : g(x) \leq 0, h(x) = 0\right\}$,

$\mathcal{F}^{\mathrm{cv}}(Z) := \left\{x \in Z : g_Z^{\mathrm{cv}}(x) \leq 0, h_Z^{\mathrm{cv}}(x) \leq 0, h_Z^{\mathrm{cc}}(x) \geq 0\right\}$.

The convex relaxation-based lower bounding scheme is said to have convergence of order $\beta > 0$ at

1. a feasible point $x \in X$ if $\exists \tau \geq 0$ s.t. $\forall Z \in \mathbb{I}X$ with $x \in Z$,

$$\min_{z \in \mathcal{F}(Z)} f(z) - \min_{z \in \mathcal{F}^{\mathrm{cv}}(Z)} f_Z^{\mathrm{cv}}(z) \leq \tau w(Z)^{\beta}.$$

"The lower bound has to converge to the minimum objective value with order at least β"

2. an infeasible point $x \in X$ if $\exists \bar{\tau} \geq 0$ s.t. $\forall Z \in \mathbb{I}X$ with $x \in Z$,

$$d(\bar{g}(Z), \mathbb{R}_-^{m_I}) - d(\bar{g}_Z^{\mathrm{cv}}(Z), \mathbb{R}_-^{m_I}) \leq \bar{\tau} w(Z)^{\beta}, \text{ and}$$

$$d(\bar{h}(Z), \{0\}) - d(I_E(Z), \{0\}) \leq \bar{\tau} w(Z)^{\beta},$$

where $(I_E(Z))\big|_{Z \in \mathbb{I}X}$ is defined as

$$(I_E(Z))\big|_{Z \in \mathbb{I}X} := \left(\left\{w \in \mathbb{R}^{m_E} : h_Z^{\mathrm{cv}}(x) \leq w \leq h_Z^{\mathrm{cc}}(x) \text{ for some } x \in Z\right\}\right)_{Z \in \mathbb{I}X}.$$

"The image of constraint relaxations has to converge (in distance) to the image of the true constraints with order at least β"

Conditions for first-order convergence

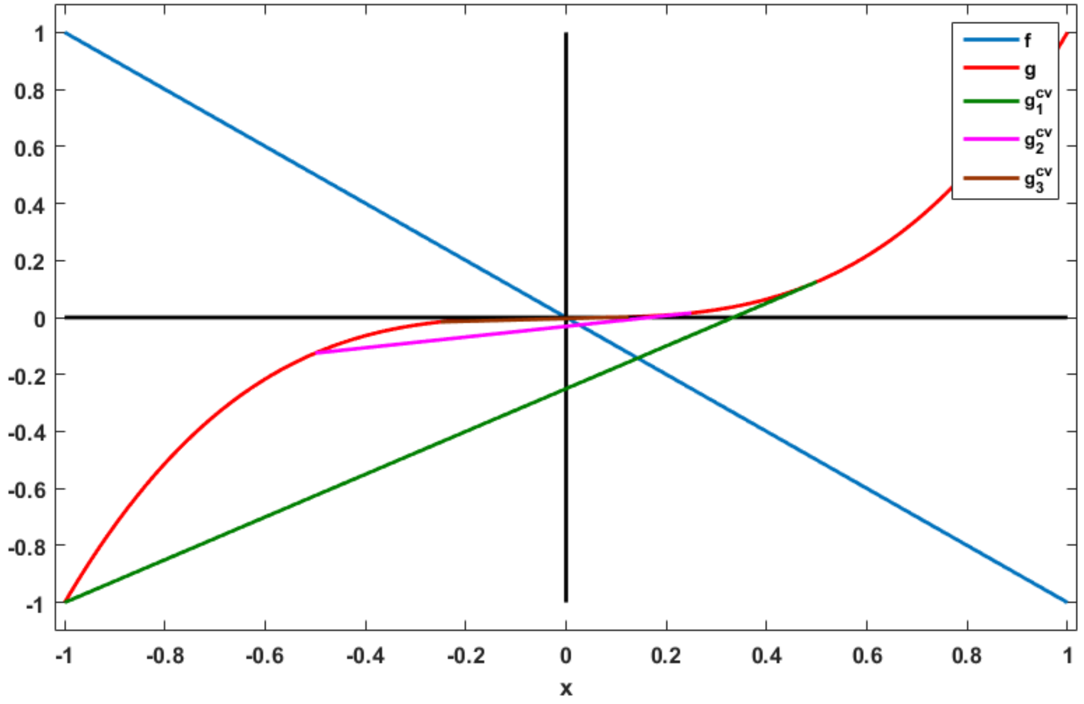
- Sufficient conditions for first-order convergence

Theorem: Suppose

1. $f, g_j, j = 1, \cdots, m_I,$ and $h_k, k = 1, \cdots, m_E,$ are Lipschitz continuous on X.
2. The schemes $(f_Z^{\text{cv}})\big|_{Z \in \mathbb{I}X}, (g_{j,Z}^{\text{cv}})\big|_{Z \in \mathbb{I}X}, j = 1, \cdots, m_I,$ and $(h_{k,Z}^{\text{cv}}, h_{k,Z}^{\text{cc}})\big|_{Z \in \mathbb{I}X}, k = 1, \cdots, m_E,$ are at least first-order pointwise convergent on X.

Then, the convex relaxation-based lower bounding scheme is at least first-order convergent on X.

$$
\begin{aligned}
&\min_x \quad -x \\
&\text{s.t.} \quad x^3 \le 0, \\
&\qquad\quad x \in [-1, 1].
\end{aligned}
$$



Conditions for second-order convergence

- Sufficient conditions for second-order convergence

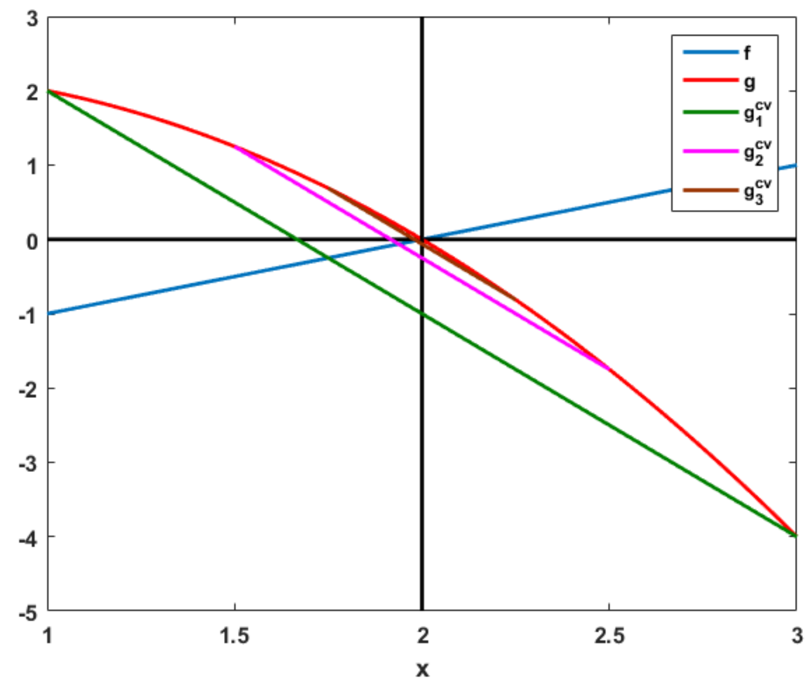
Theorem: Suppose

1. $f, g_j, j = 1, \cdots, m_I$, and $h_k, k = 1, \cdots, m_E$, are $\mathcal{C}^2$ on X.
2. The schemes $(f_Z^{\text{cv}})\big|_{Z \in \mathbb{I}X}, (g_{j,Z}^{\text{cv}})\big|_{Z \in \mathbb{I}X}, j = 1, \cdots, m_I$, and $(h_{k,Z}^{\text{cv}}, h_{k,Z}^{\text{cc}})\big|_{Z \in \mathbb{I}X}, k = 1, \cdots, m_E$,
   are at least second-order pointwise convergent on X.

Then, the convex relaxation-based lower bounding scheme is at least second-order convergent at

1. $x \in X$ for which $\exists (\mu, \lambda) \in \mathbb{R}_+^{m_I} \times \mathbb{R}^{m_E}$ such that (x, μ, λ) is a KKT point
2. $x \in X$ with $g(x) < 0$ (when $m_E = 0$)
3. infeasible $x \in X$

$$\begin{aligned} \min_x \ & x \\ \text{s.t.} \ & -x^2 + x + 2 \le 0, \\ & x \in [1, 3]. \end{aligned}$$



Clustering in Constrained Global Optimization

Consider the problem

$$\min_{x\in X} f(x)$$
$$\text{s.t. } g(x) \le 0,$$
$$h(x) = 0,$$

where $X \subset \mathbb{R}^n$ is a nonempty open bounded convex set, $f : X \to \mathbb{R}$, $g : X \to \mathbb{R}^{m_I}$, $h : X \to \mathbb{R}^{m_E}$.

Assume

1. $f, g,$ and h are $\mathcal{C}^2$ on X
2. The constraints define a compact set inside X
3. $x^* \in X$ is a global minimum of the above problem, and the B&B algorithm has found the upper bound $UBD = f(x^*)$ early on
4. The termination tolerance $\varepsilon \ll 1$ and the algorithm fathoms node k when $UBD - LBD_k \le \varepsilon$

Clustering in Constrained Global Optimization

Suppose the lower bounding scheme

1. has convergence of order $\beta^{*} > 0$ at feasible points with a prefactor $\tau^{*} > 0$
2. has convergence of order $\beta^{\mathrm{I}} > 0$ at infeasible points with a prefactor $\tau^{\mathrm{I}} > 0$

Clustering in Constrained Global Optimization

Suppose the lower bounding scheme

1. has convergence of order $\beta^{*} > 0$ at feasible points with a prefactor $\tau^{*} > 0$
2. has convergence of order $\beta^{\mathrm{I}} > 0$ at infeasible points with a prefactor $\tau^{\mathrm{I}} > 0$

Suppose $(f_Z^{\mathrm{cv}})\big|_{Z \in \mathbb{I}X}$ has convergence of order $\beta^{\mathrm{f}} > 0$ on X with a prefactor $\tau^{\mathrm{f}} > 0$

Let $\varepsilon^{f}, \varepsilon^{o}$ be such that $\left(\dfrac{\varepsilon^{\mathrm{f}}}{\tau^{\mathrm{I}}}\right)^{\frac{1}{\beta^{\mathrm{I}}}} = \left(\dfrac{\varepsilon^{\mathrm{o}}}{\tau^{\mathrm{f}}}\right)^{\frac{1}{\beta^{\mathrm{f}}}} = \left(\dfrac{\varepsilon}{\tau^{*}}\right)^{\frac{1}{\beta^{*}}}.$

Clustering in Constrained Global Optimization

Suppose the lower bounding scheme

1. has convergence of order $\beta^{*} > 0$ at feasible points with a prefactor $\tau^{*} > 0$
2. has convergence of order $\beta^{\mathrm{I}} > 0$ at infeasible points with a prefactor $\tau^{\mathrm{I}} > 0$

Suppose $(f_Z^{\mathrm{cv}})\big|_{Z \in \mathbb{I}X}$ has convergence of order $\beta^{\mathrm{f}} > 0$ on X with a prefactor $\tau^{\mathrm{f}} > 0$

Let $\varepsilon^{f}, \varepsilon^{o}$ be such that $\left(\dfrac{\varepsilon^{\mathrm{f}}}{\tau^{\mathrm{I}}}\right)^{\frac{1}{\beta^{\mathrm{I}}}} = \left(\dfrac{\varepsilon^{\mathrm{o}}}{\tau^{\mathrm{f}}}\right)^{\frac{1}{\beta^{\mathrm{f}}}} = \left(\dfrac{\varepsilon}{\tau^{*}}\right)^{\frac{1}{\beta^{*}}}$.

Partition X into regions $X_1, \cdots, X_5$ such that

$$X_1 = \left\{x \in X : \max\left\{d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} > \varepsilon^{\mathrm{f}}\right\},$$

$$X_2 = \left\{x \in X : \max\left\{d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} \in (0, \varepsilon^{\mathrm{f}}] \text{ and } f(x) - f(x^*) > \varepsilon^{\mathrm{o}}\right\},$$

$$X_3 = \left\{x \in X : \max\left\{d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} \in (0, \varepsilon^{\mathrm{f}}] \text{ and } f(x) - f(x^*) \le \varepsilon^{\mathrm{o}}\right\},$$

$$X_4 = \left\{x \in X : \max\left\{d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} = 0 \text{ and } f(x) - f(x^*) > \varepsilon\right\},$$

$$X_5 = \left\{x \in X : \max\left\{d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} = 0 \text{ and } f(x) - f(x^*) \le \varepsilon\right\}.$$

Clustering in Constrained Global Optimization

Partition X into regions $X_1, \cdots, X_5$ such that

$X_1 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_{-}^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} > \varepsilon^{\mathrm{f}} \right\},$ "quite infeasible"

$X_2 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_{-}^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} \in (0, \varepsilon^{\mathrm{f}}] \text{ and } f(x) - f(x^*) > \varepsilon^{\mathrm{o}} \right\},$ "nearly feasible" but have "poor objective value"

$X_3 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_{-}^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} \in (0, \varepsilon^{\mathrm{f}}] \text{ and } f(x) - f(x^*) \leq \varepsilon^{\mathrm{o}} \right\},$ "nearly feasible" and have "good objective value"

$X_4 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_{-}^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} = 0 \text{ and } f(x) - f(x^*) > \varepsilon \right\},$ feasible but "quite suboptimal"

$X_5 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_{-}^{m_I}\right), d\left(\{h(x)\}, \{0\}\right)\right\} = 0 \text{ and } f(x) - f(x^*) \leq \varepsilon \right\}.$ feasible and "nearly optimal"

Clustering in Constrained Global Optimization

Partition X into regions $X_1, \cdots, X_5$ such that

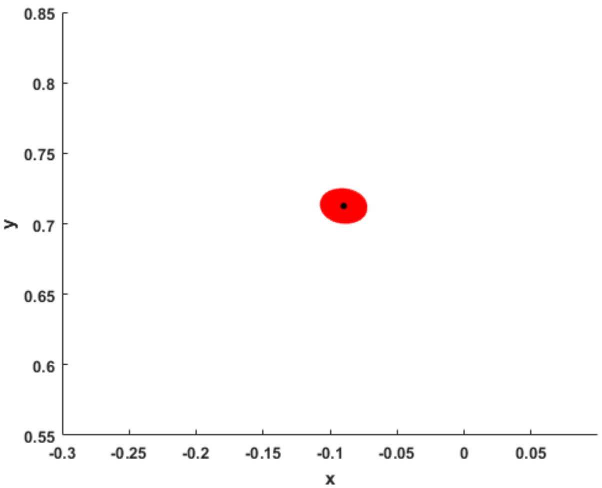
$X_1 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right) \right\} > \varepsilon^{\mathrm{f}} \right\},$ "quite infeasible"

$X_2 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right) \right\} \in (0, \varepsilon^{\mathrm{f}}] \text{ and } f(x) - f(x^*) > \varepsilon^{\mathrm{o}} \right\},$ "nearly feasible" but have "poor objective value"

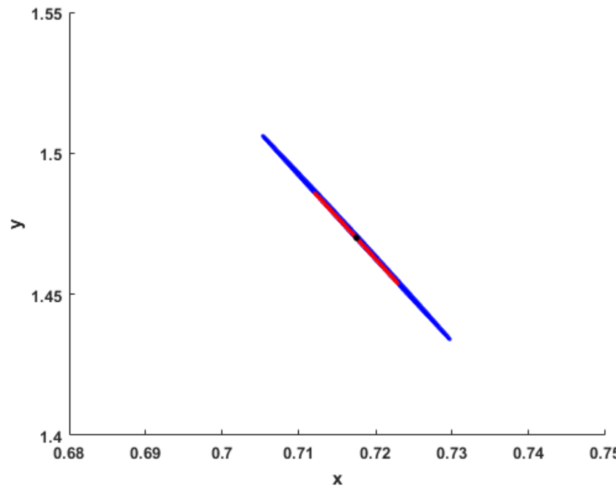
$X_3 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right) \right\} \in (0, \varepsilon^{\mathrm{f}}] \text{ and } f(x) - f(x^*) \le \varepsilon^{\mathrm{o}} \right\},$ "nearly feasible" and have "good objective value"

$X_4 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right) \right\} = 0 \text{ and } f(x) - f(x^*) > \varepsilon \right\},$ feasible but "quite suboptimal"

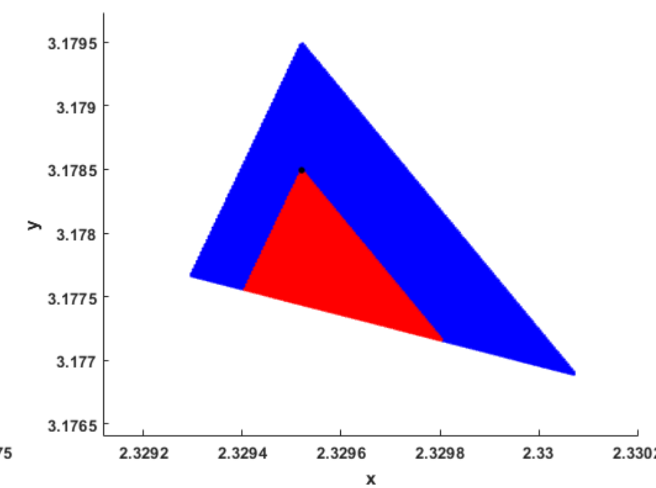
$X_5 = \left\{ x \in X : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I}\right), d\left(\{h(x)\}, \{0\}\right) \right\} = 0 \text{ and } f(x) - f(x^*) \le \varepsilon \right\}.$ feasible and "nearly optimal"



Unconstrained



Equality constrained



Inequality constrained

More Definitions

- Neighborhood of a point

  Let $x \in X \subset \mathbb{R}^n$. For any $\alpha > 0$, $p \in \mathbb{N}$, the set $\mathcal{N}_\alpha^p(x) = \left\{ z \in X : \|z - x\|_p < \alpha \right\}$ is called the $\alpha-$neighborhood of x in X with respect to the $p-$norm.

- Strict local minimum

  A point $\bar{x} \in \mathcal{F}(X)$ is called a strict local minimum if $\bar{x}$ is a local minimum and $\exists \alpha > 0$ such that $f(x) > f(\bar{x})$, $\forall x \in \mathcal{N}_\alpha^2(\bar{x}) \cap \mathcal{F}(X)$ s.t. $x \neq \bar{x}$.

- Nonisolated feasible point

  A feasible point $\bar{x} \in \mathcal{F}(X)$ is said to be nonisolated if $\forall \alpha > 0$, $\exists z \in \mathcal{N}_\alpha^2(\bar{x}) \cap \mathcal{F}(X)$ s.t. $z \neq \bar{x}$.

- Set of active inequality constraints

  Let $x \in \mathcal{F}(X)$ be a feasible point. The set of active inequality constraints at x is given by

$$\mathcal{A}(x) = \left\{ j \in \{1, \cdots, m_I\} : g_j(x) = 0 \right\}.$$

First-Order Clustering in X_5

$X_3 = \{x \in X \text{ which are "nearly feasible" and have "good objective value"}\}$,

$X_5 = \{x \in X \text{ which are feasible and "nearly optimal"}\}$.

First-Order Clustering in X_5

$X_3 = \{x \in X \text{ which are "nearly feasible" and have "good objective value"}\}$,

$X_5 = \{x \in X \text{ which are feasible and "nearly optimal"}\}$.

Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0$ s.t.

$$L = \inf_{\{d : \|d\| = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}(X)\}} \nabla f(x^*)^{\mathrm{T}} d > 0.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L \left\| x - x^* \right\|_1 \le 2\varepsilon \right\}.$$

First-Order Clustering in X_5

$X_3 = \{x \in X$ which are "nearly feasible" and have "good objective value"$\}$,

$X_5 = \{x \in X$ which are feasible and "nearly optimal"$\}$.

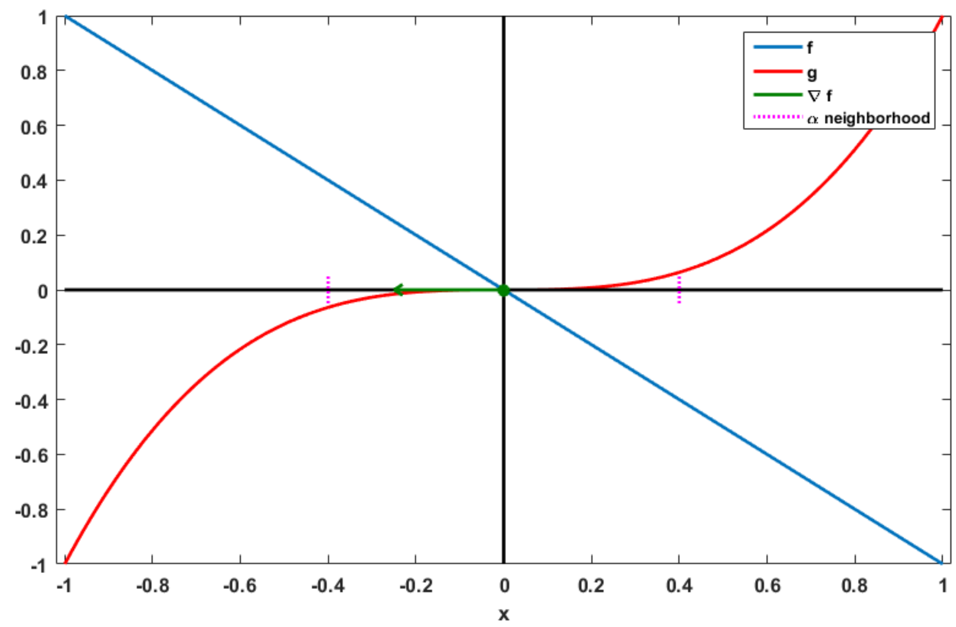
Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0$ s.t.

$$L = \inf_{\{d : \|d\|=1, \exists t>0 \text{ s.t. } (x^*+td) \in \mathcal{N}^1_\alpha(x^*) \cap \mathcal{F}(X)\}} \nabla f(x^*)^\mathrm{T} d > 0.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}^1_{\hat{\alpha}}(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}^1_{\hat{\alpha}}(x^*) : L \left\| x - x^* \right\|_1 \le 2\varepsilon \right\}.$$

$$\begin{aligned} \min_x \quad & -x \\ \text{s.t.} \quad & x^3 \le 0, \\ & x \in [-1, 1]. \end{aligned}$$



First-Order Clustering in X_5

$X_3 = \{x \in X \text{ which are "nearly feasible" and have "good objective value"}\}$,

$X_5 = \{x \in X \text{ which are feasible and "nearly optimal"}\}$.

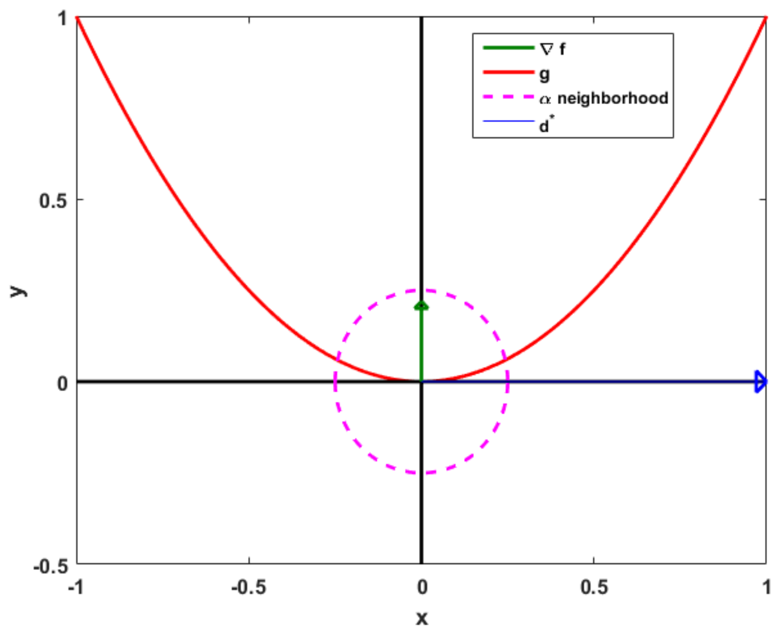
Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0$ s.t.

$$L = \inf_{\{d: \|d\|=1, \exists t>0 \text{ s.t. } (x^*+td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}(X)\}} \nabla f(x^*)^{\mathrm{T}} d > 0.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L \left\| x - x^* \right\|_1 \le 2\varepsilon \right\}.$$

$$\min_{x,y} \ y$$
$$\text{s.t. } x^2 - y \le 0,$$
$$x \in [-1,1], y \in [-1,1].$$



First-Order Clustering in X_5

Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0$ s.t.

$$L = \inf_{\left\{d : \|d\|=1, \exists t>0 \text{ s.t. } (x^*+td) \in \mathcal{N}_{\alpha}^{1}(x^*) \cap \mathcal{F}(X)\right\}} \nabla f(x^*)^{\mathrm{T}} d > 0.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^{1}(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^{1}(x^*) : L \left\| x - x^* \right\|_1 \le 2\varepsilon \right\}.$$

Theorem: Let $\delta = \left(\dfrac{\varepsilon}{\tau^*} \right)^{\frac{1}{\beta^*}}$, $r = \dfrac{2\varepsilon}{L}$.

1. If $\delta \ge 2r$, then let $N = 1$.

2. If $\dfrac{2r}{m-1} > \delta \ge \dfrac{2r}{m}$ for some $m \in \mathbb{N}$ with $m \le n$, $2 \le m \le 6$, then let

$$N = \sum_{i=0}^{m-1} 2^i \binom{n}{i} + 2n \left\lceil \frac{m-3}{3} \right\rceil.$$

3. Otherwise, let

$$N = \left\lceil 2 \tau^{*\frac{1}{\beta^*}} \varepsilon^{1-\frac{1}{\beta^*}} L^{-1} \right\rceil^{n-1} \left(\left\lceil 2 \tau^{*\frac{1}{\beta^*}} \varepsilon^{1-\frac{1}{\beta^*}} L^{-1} \right\rceil + 2n \left\lceil \tau^{*\frac{1}{\beta^*}} \varepsilon^{1-\frac{1}{\beta^*}} L^{-1} \right\rceil \right).$$

N is an upper bound on the number of boxes of width δ required to cover $\hat{X}_5$.

First-Order Clustering in X_5

Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0$ s.t.

$$L = \inf_{\{d:\|d\|=1, \exists t>0 \text{ s.t. } (x^*+td)\in\mathcal{N}_\alpha^1(x^*)\cap\mathcal{F}(X)\}} \nabla f(x^*)^{\mathrm{T}} d > 0.$$

Then $\exists \hat{\alpha} \in (0,\alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L\left\| x - x^* \right\|_1 \leq 2\varepsilon \right\}.$$

Number of boxes required to cover $\hat{X}_5$ when $\beta^* = 1$

Case	Number of boxes
$\tau^* \leq \frac{L}{4}$	1
$\frac{L}{4} < \tau^* \leq \frac{2L}{4}$	$1 + 2n$
$\frac{2L}{4} < \tau^* \leq \frac{3L}{4}$	$1 + 2n^2$
$\frac{3L}{4} < \tau^* \leq \frac{4L}{4}$	$1 + \frac{14}{3}n - 2n^2 + \frac{4}{3}n^3$
$\vdots$	$\vdots$
$\frac{6L}{4} < \tau^*$	$\lceil 2\tau^* L^{-1} \rceil^{n-1} \left(\lceil 2\tau^* L^{-1} \rceil + 2n \lceil \tau^* L^{-1} \rceil \right)$

Second-Order Clustering in X_5

Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0, \gamma > 0$ s.t.

$$\nabla f(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x^*) d \geq \gamma d^{\mathrm{T}} d, \quad \forall d \in \left\{ d : (x^* + d) \in \mathcal{N}_{\alpha}^2(x^*) \cap \mathcal{F}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \gamma \left\| x - x^* \right\|_2^2 \leq 2\varepsilon \right\}.$$

Second-Order Clustering in X_5

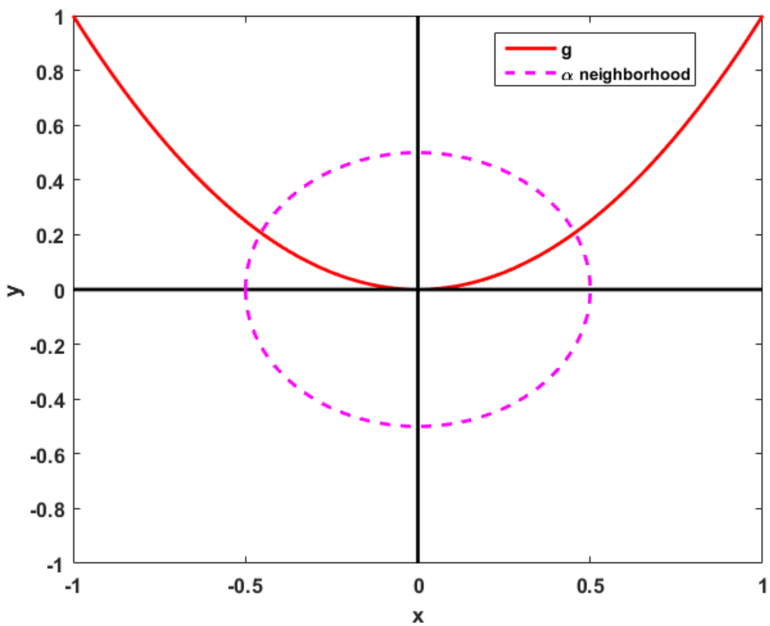
Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0, \gamma > 0$ s.t.

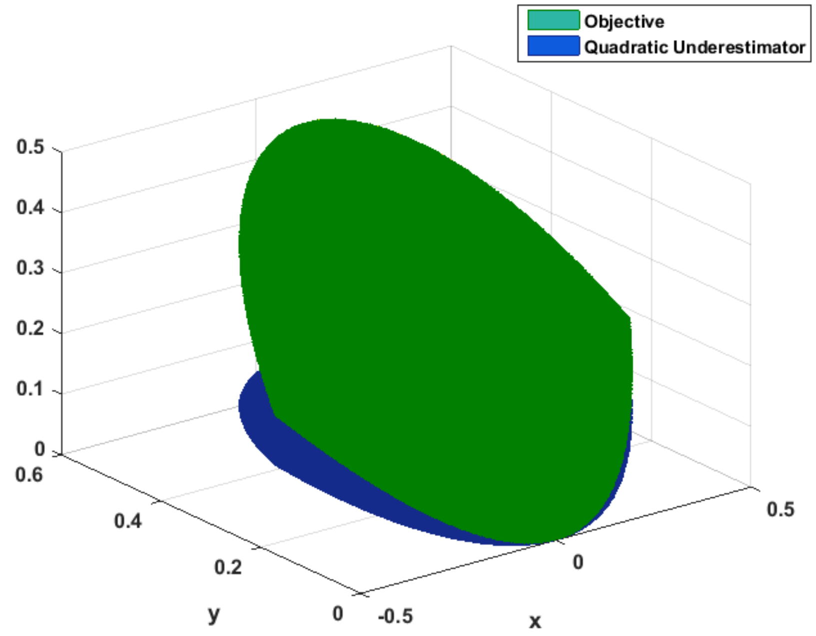
$$\nabla f(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x^*) d \geq \gamma d^{\mathrm{T}} d, \quad \forall d \in \left\{ d : (x^* + d) \in \mathcal{N}_\alpha^2(x^*) \cap \mathcal{F}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \gamma \left\| x - x^* \right\|_2^2 \leq 2\varepsilon \right\}.$$

$$\begin{aligned} \min_{x,y} \ & y \\ \text{s.t. } \ & x^2 - y \leq 0, \\ & x \in [-1,1], y \in [-1,1]. \end{aligned}$$





Second-Order Clustering in X_5

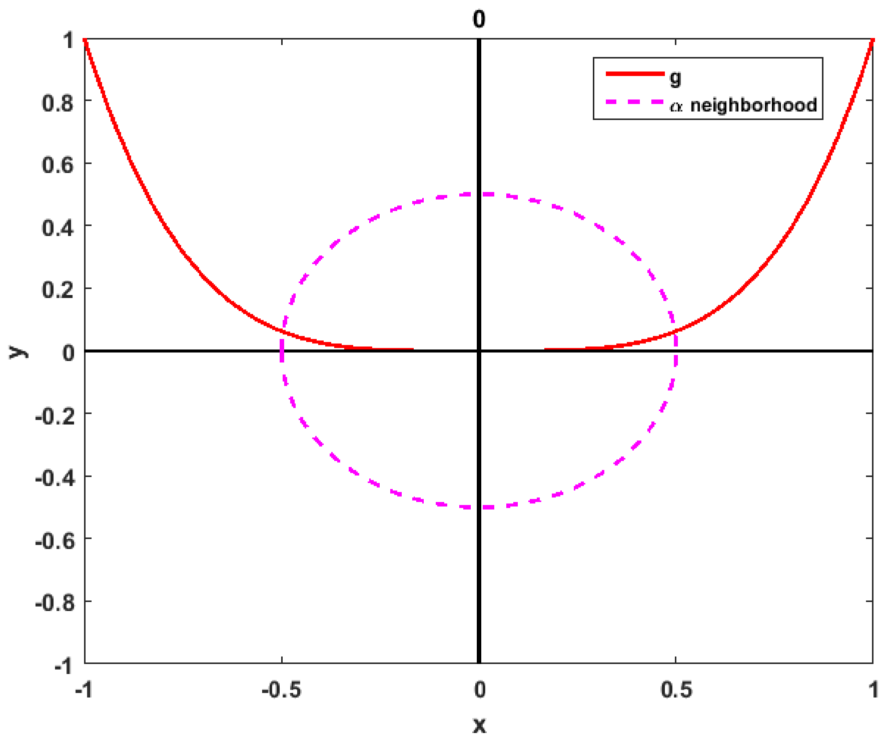
Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0, \gamma > 0$ s.t.

$$\nabla f(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x^*) d \geq \gamma d^{\mathrm{T}} d, \quad \forall d \in \left\{ d : (x^* + d) \in \mathcal{N}_\alpha^2(x^*) \cap \mathcal{F}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \gamma \left\| x - x^* \right\|_2^2 \leq 2\varepsilon \right\}.$$

$$\begin{aligned} \min_{x,y} \; & y \\ \text{s.t. } \; & x^4 - y \leq 0, \\ & x \in [-1,1], y \in [-1,1]. \end{aligned}$$



Second-Order Clustering in X_5

Lemma: Suppose x^* is a nonisolated feasible point and $\exists \alpha > 0, \gamma > 0$ s.t.

$$\nabla f(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x^*) d \geq \gamma d^{\mathrm{T}} d, \quad \forall d \in \left\{ d : (x^* + d) \in \mathcal{N}_\alpha^2(x^*) \cap \mathcal{F}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region $\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_5$ is overestimated by

$$\hat{X}_5 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \gamma \left\| x - x^* \right\|_2^2 \leq 2\varepsilon \right\}.$$

Number of boxes required to cover $\hat{X}_5$ when $\beta^* = 2$

Case	Number of boxes
$\tau^* \leq \frac{\gamma}{8}$	1
$\frac{\gamma}{8} < \tau^* \leq \frac{2\gamma}{8}$	$1 + 2n$
$\frac{2\gamma}{8} < \tau^* \leq \frac{3\gamma}{8}$	$1 + 2n^2$
$\frac{3\gamma}{8} < \tau^* \leq \frac{4\gamma}{8}$	$1 + \frac{8}{3}n - 2n^2 + \frac{4}{3}n^3$
$\vdots$	$\vdots$
$\frac{18\gamma}{8} < \tau^*$	$\left\lceil 2\sqrt{\tau^* \gamma^{-1}} \right\rceil^{n-1} \left(\left\lceil 2\sqrt{\tau^* \gamma^{-1}} \right\rceil + 2n \left\lceil (\sqrt{2} - 1)\sqrt{\tau^* \gamma^{-1}} \right\rceil \right)$

First-Order Clustering in X_3

$X_3 = \{x \in X \text{ which are "nearly feasible" and have "good objective value"}\},$

$X_5 = \{x \in X \text{ which are feasible and "nearly optimal"}\}.$

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0$ and a set $\mathcal{D}_1$ s.t.

$$L_f = \inf_{d \in \mathcal{D}_1 \cap \mathcal{D}_I} \nabla f(x^*)^{\mathrm{T}} d > 0,$$

$$L_I = \inf_{d \in \mathcal{D}_I \setminus \mathcal{D}_1} \max\left\{ \max_{j \in \mathcal{A}(x^*)} \left\{\nabla g_j(x^*)^{\mathrm{T}} d\right\}, \max_{k \in \{1,\cdots,m_E\}} \left\{\left|\nabla h_k(x^*)^{\mathrm{T}} d\right|\right\}\right\} > 0,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \left\{d : \|d\|_1 = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X)\right\}.$$

First-Order Clustering in X_3

$X_3 = \{x \in X \text{ which are "nearly feasible" and have "good objective value"}\}$,

$X_5 = \{x \in X \text{ which are feasible and "nearly optimal"}\}$.

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0$ and a set $\mathcal{D}_1$ s.t.

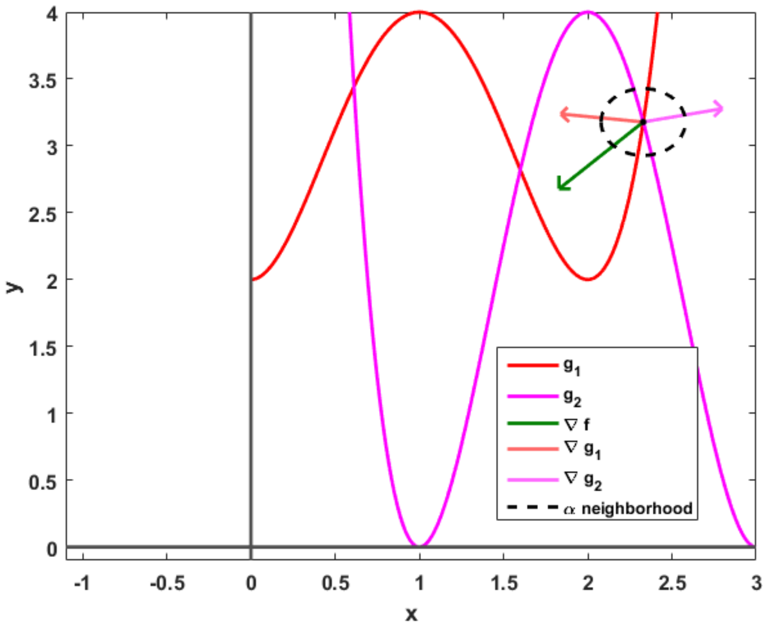
$$L_f = \inf_{d \in \mathcal{D}_1 \cap \mathcal{D}_I} \nabla f(x^*)^{\mathrm{T}} d > 0,$$

$$L_I = \inf_{d \in \mathcal{D}_I \setminus \mathcal{D}_1} \max\left\{ \max_{j \in \mathcal{A}(x^*)} \left\{ \nabla g_j(x^*)^{\mathrm{T}} d \right\}, \max_{k \in \{1,\cdots,m_E\}} \left\{ \left| \nabla h_k(x^*)^{\mathrm{T}} d \right| \right\} \right\} > 0,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \left\{ d : \|d\|_1 = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) \right\}.$$

$$\begin{aligned} \min_{x,y} \quad & -x - y \\ \text{s.t.} \quad & y \le 2 + 2x^4 - 8x^3 + 8x^2, \\ & y \le 4x^4 - 32x^3 + 88x^2 - 96x + 36, \\ & x \in [0,3],\ y \in [0,4]. \end{aligned}$$



First-Order Clustering in X_3

$X_3 = \{x \in X \text{ which are "nearly feasible" and have "good objective value"}\}$,

$X_5 = \{x \in X \text{ which are feasible and "nearly optimal"}\}$.

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0$ and a set $\mathcal{D}_1$ s.t.

$$L_f = \inf_{d \in \mathcal{D}_1 \cap \mathcal{D}_I} \nabla f(x^*)^{\mathrm{T}} d > 0,$$

$$L_I = \inf_{d \in \mathcal{D}_I \setminus \mathcal{D}_1} \max\left\{ \max_{j \in \mathcal{A}(x^*)} \{\nabla g_j(x^*)^{\mathrm{T}} d\}, \max_{k \in \{1,\cdots,m_E\}} \{|\nabla h_k(x^*)^{\mathrm{T}} d|\} \right\} > 0,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \{d : \|d\|_1 = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X)\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \{x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_1 \cap \mathcal{D}_I, t > 0\}$$

is overestimated by

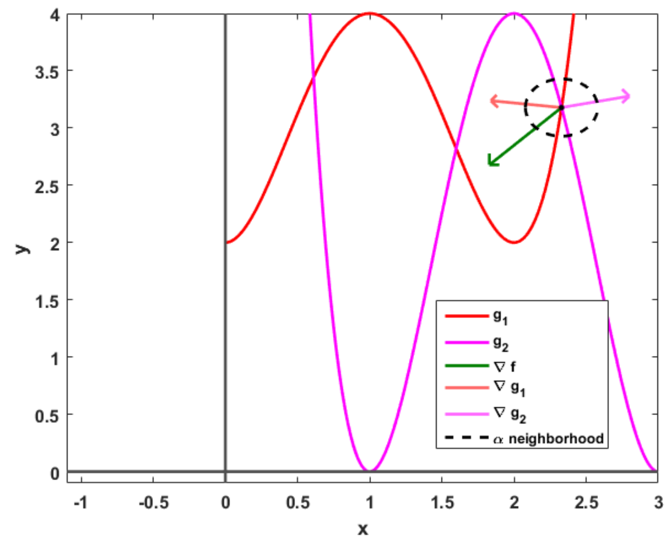
$$\hat{X}_3^1 = \{x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_f \|x - x^*\|_1 \le 2\varepsilon^{\mathrm{o}}\}$$

and the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \{x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1, t > 0\}$$

is overestimated by

$$\hat{X}_3^2 = \{x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_I \|x - x^*\|_1 \le 2\varepsilon^{\mathrm{f}}\}.$$



First-Order Clustering in X_3

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0$ and a set $\mathcal{D}_1$ s.t.

$$L_f = \inf_{d \in \mathcal{D}_1 \cap \mathcal{D}_I} \nabla f(x^*)^{\mathrm{T}} d > 0,$$

$$L_I = \inf_{d \in \mathcal{D}_I \setminus \mathcal{D}_1} \max\left\{ \max_{j \in \mathcal{A}(x^*)} \left\{ \nabla g_j(x^*)^{\mathrm{T}} d \right\}, \max_{k \in \{1,\cdots,m_E\}} \left\{ \left| \nabla h_k(x^*)^{\mathrm{T}} d \right| \right\} \right\} > 0,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \left\{ d : \|d\|_1 = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_1 \cap \mathcal{D}_I, t > 0 \right\}$$

is overestimated by

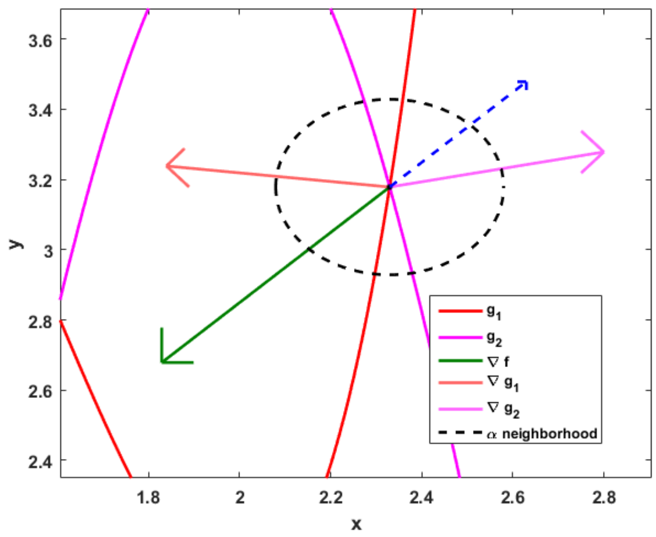
$$\hat{X}_3^1 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_f \left\| x - x^* \right\|_1 \le 2\varepsilon^{\mathrm{o}} \right\}$$

and the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1, t > 0 \right\}$$

is overestimated by

$$\hat{X}_3^2 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_I \left\| x - x^* \right\|_1 \le 2\varepsilon^{\mathrm{f}} \right\}.$$



First-Order Clustering in X_3

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0$ and a set $\mathcal{D}_1$ s.t.

$$L_f = \inf_{d \in \mathcal{D}_1 \cap \mathcal{D}_I} \nabla f(x^*)^{\mathrm{T}} d > 0,$$

$$L_I = \inf_{d \in \mathcal{D}_I \setminus \mathcal{D}_1} \max\left\{ \max_{j \in \mathcal{A}(x^*)} \left\{ \nabla g_j(x^*)^{\mathrm{T}} d \right\}, \max_{k \in \{1,\cdots,m_E\}} \left\{ \left| \nabla h_k(x^*)^{\mathrm{T}} d \right| \right\} \right\} > 0,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \left\{ d : \|d\|_1 = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_1 \cap \mathcal{D}_I, t > 0 \right\}$$

is overestimated by

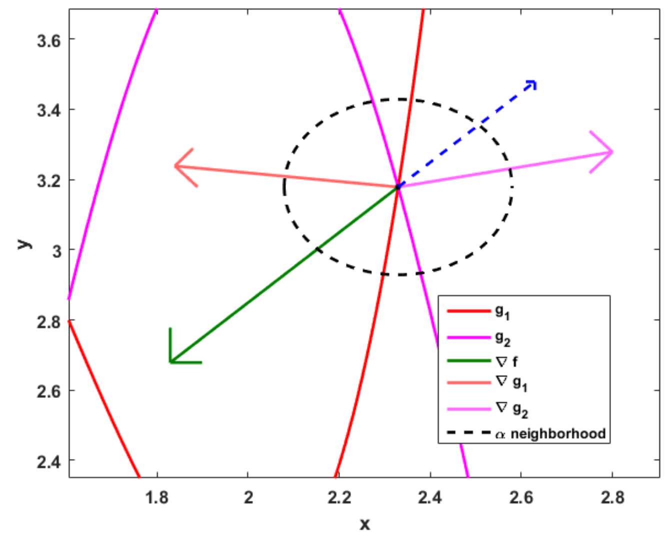
$$\hat{X}_3^1 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_f \left\| x - x^* \right\|_1 \le 2\varepsilon^{\mathrm{o}} \right\}$$

and the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1, t > 0 \right\}$$

is overestimated by

$$\hat{X}_3^2 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_I \left\| x - x^* \right\|_1 \le 2\varepsilon^{\mathrm{f}} \right\}.$$



Furthermore, suppose x^* is at the center of a box, B_δ, of width $\delta = \left(\frac{\varepsilon}{\tau^*} \right)^{\frac{1}{\beta^*}}$ placed while covering $\hat{X}_5$.

Then the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1, t > 0 \right\} \setminus B_\delta$$

is overestimated by

$$\left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I} \right), d\left(\{h(x)\}, \{0\} \right) \right\} \in \left[\frac{L_I}{4}\delta, \varepsilon^{\mathrm{f}} \right] \right\}$$

First-Order Clustering in X_3

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0$ and a set $\mathcal{D}_1$ s.t.

$$L_f = \inf_{d \in \mathcal{D}_1 \cap \mathcal{D}_I} \nabla f(x^*)^{\mathrm{T}} d > 0,$$

$$L_I = \inf_{d \in \mathcal{D}_I \setminus \mathcal{D}_1} \max\left\{ \max_{j \in \mathcal{A}(x^*)} \left\{ \nabla g_j(x^*)^{\mathrm{T}} d \right\}, \max_{k \in \{1,\cdots,m_E\}} \left\{ \left| \nabla h_k(x^*)^{\mathrm{T}} d \right| \right\} \right\} > 0,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \left\{ d : \|d\|_1 = 1, \exists t > 0 \text{ s.t. } (x^* + td) \in \mathcal{N}_\alpha^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_1 \cap \mathcal{D}_I, t > 0 \right\}$$

is overestimated by

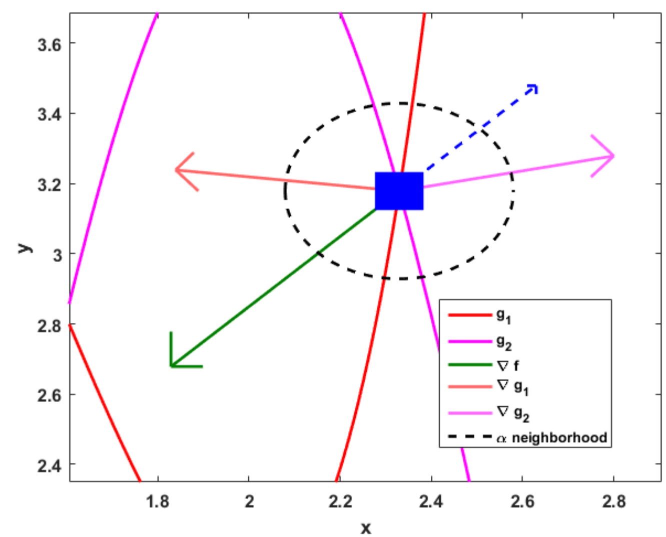
$$\hat{X}_3^1 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_f \left\| x - x^* \right\|_1 \le 2\varepsilon^{\mathrm{o}} \right\}$$

and the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1, t > 0 \right\}$$

is overestimated by

$$\hat{X}_3^2 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : L_I \left\| x - x^* \right\|_1 \le 2\varepsilon^{\mathrm{f}} \right\}.$$



Furthermore, suppose x^* is at the center of a box, B_δ, of width $\delta = \left(\frac{\varepsilon}{\tau^*} \right)^{\frac{1}{\beta^*}}$ placed while covering $\hat{X}_5$.

Then the region

$$\mathcal{N}_{\hat{\alpha}}^1(x^*) \cap X_3 \cap \left\{ x = (x^* + td) \in \mathcal{N}_{\hat{\alpha}}^1(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1, t > 0 \right\} \setminus B_\delta$$

is overestimated by

$$\left\{ x \in \mathcal{N}_{\hat{\alpha}}^1(x^*) : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I} \right), d\left(\{h(x)\}, \{0\} \right) \right\} \in \left[\frac{L_I}{4}\delta, \varepsilon^{\mathrm{f}} \right] \right\}$$

First-Order Clustering in X_3

Theorem: Suppose the conditions of the Lemma hold.

Let $\delta = \delta_f = \left(\frac{\varepsilon}{\tau^*}\right)^{\frac{1}{\beta^*}} = \left(\frac{\varepsilon^{\mathrm{o}}}{\tau^{\mathrm{f}}}\right)^{\frac{1}{\beta^{\mathrm{f}}}}, \delta_I = \left(\frac{L_I \delta}{4\tau^I}\right)^{\frac{1}{\beta^{\mathrm{I}}}} = \left(\frac{L_I}{4\tau^{\mathrm{I}}}\right)^{\frac{1}{\beta^{\mathrm{I}}}} \left(\frac{\varepsilon^{\mathrm{f}}}{\tau^{\mathrm{I}}}\right)^{\frac{1}{\left(\beta^{\mathrm{I}}\right)^2}}, \; r_I = \frac{2\varepsilon^{\mathrm{f}}}{L_I}, \; r_f = \frac{2\varepsilon^{\mathrm{o}}}{L_f}.$

1. If $\delta_I \geq 2r_I$, then let $N_I = 1$.

2. If $\frac{2r_I}{m_I^{'}-1} > \delta_I \geq \frac{2r_I}{m_I^{'}}$ for some $m_I \in \mathbb{N}$ with $m_I^{'} \leq n$, $2 \leq m_I^{'} \leq 6$, then let $N_I = \sum_{i=0}^{m_I^{'}-1} 2^i \binom{n}{i} + 2n \left\lceil \frac{m_I^{'}-3}{3} \right\rceil$.

3. Otherwise, $N_I = \left\lceil 0.5\left(\tau^{\mathrm{I}}\right)^{\frac{2}{\beta^{\mathrm{I}}}} \left(\varepsilon^{\mathrm{f}}\right)^{1-\frac{1}{\left(\beta^{\mathrm{I}}\right)^2}} L_I^{-\frac{2}{\beta^{\mathrm{I}}}} \right\rceil^{n-1} \left(\left\lceil 0.5\left(\tau^{\mathrm{I}}\right)^{\frac{2}{\beta^{\mathrm{I}}}} \left(\varepsilon^{\mathrm{f}}\right)^{1-\frac{1}{\left(\beta^{\mathrm{I}}\right)^2}} L_I^{-\frac{2}{\beta^{\mathrm{I}}}} \right\rceil + 2n \left\lceil 0.25\left(\tau^{\mathrm{I}}\right)^{\frac{2}{\beta^{\mathrm{I}}}} \left(\varepsilon^{\mathrm{f}}\right)^{1-\frac{1}{\left(\beta^{\mathrm{I}}\right)^2}} L_I^{-\frac{2}{\beta^{\mathrm{I}}}} \right\rceil \right).$

4. If $\delta_f \geq 2r_f$, then let $N_f = 1$.

5. If $\frac{2r_f}{m_f-1} > \delta_f \geq \frac{2r_f}{m_f}$ for some $m_f \in \mathbb{N}$ with $m_f \leq n$, $2 \leq m_f \leq 6$, then let $N_f = \sum_{i=0}^{m_f-1} 2^i \binom{n}{i} + 2n \left\lceil \frac{m_f-3}{3} \right\rceil$.

6. Otherwise, $N_f = \left\lceil 2\left(\tau^{\mathrm{f}}\right)^{\frac{1}{\beta^{\mathrm{f}}}} \left(\varepsilon^{\mathrm{o}}\right)^{1-\frac{1}{\beta^{\mathrm{f}}}} L_f^{-\frac{1}{\beta^{\mathrm{f}}}} \right\rceil^{n-1} \left(\left\lceil 2\left(\tau^{\mathrm{f}}\right)^{\frac{1}{\beta^{\mathrm{f}}}} \left(\varepsilon^{\mathrm{o}}\right)^{1-\frac{1}{\beta^{\mathrm{f}}}} L_f^{-\frac{1}{\beta^{\mathrm{f}}}} \right\rceil + 2n \left\lceil \left(\tau^{\mathrm{f}}\right)^{\frac{1}{\beta^{\mathrm{f}}}} \left(\varepsilon^{\mathrm{o}}\right)^{1-\frac{1}{\beta^{\mathrm{f}}}} L_f^{-\frac{1}{\beta^{\mathrm{f}}}} \right\rceil \right).$

N_I is an upper bound on the number of boxes of width δ_I required to cover $\hat{X}_3^2 \setminus \hat{X}_5$ and
N_f is an upper bound on the number of boxes of width δ_f required to cover $\hat{X}_3^1$.

First-Order Clustering in X_3

Number of boxes required to cover $\hat{X}_3^1$ when $\beta^f = 1$

Case	Number of boxes
$\tau^f \le \frac{L_f}{4}$	1
$\frac{L_f}{4} < \tau^f \le \frac{2L_f}{4}$	$1 + 2n$
$\frac{2L_f}{4} < \tau^f \le \frac{3L_f}{4}$	$1 + 2n^2$
$\frac{3L_f}{4} < \tau^f \le \frac{4L_f}{4}$	$1 + \frac{14}{3}n - 2n^2 + \frac{4}{3}n^3$
$\vdots$	$\vdots$
$\frac{6L_f}{4} < \tau^f$	$\lceil 2\tau^f {L_f}^{-1} \rceil^{n-1} \left(\lceil 2\tau^f {L_f}^{-1} \rceil + 2n \lceil \tau^f {L_f}^{-1} \rceil \right)$

Number of boxes required to cover $\hat{X}_3^2 \backslash \hat{X}_5$ when $\beta^I = 1$

Case	Number of boxes
$\tau^I \le \frac{L_I}{4}$	1
$\frac{L_I}{4} < \tau^I \le \frac{\sqrt{2}L_I}{4}$	$1 + 2n$
$\frac{\sqrt{2}L_I}{4} < \tau^I \le \frac{\sqrt{3}L_I}{4}$	$1 + 2n^2$
$\frac{\sqrt{3}L_I}{4} < \tau^I \le \frac{\sqrt{4}L_I}{4}$	$1 + \frac{14}{3}n - 2n^2 + \frac{4}{3}n^3$
$\vdots$	$\vdots$
$\frac{\sqrt{6}L_I}{4} < \tau^I$	$\lceil 0.5(\tau^I)^2 {L_I}^{-2} \rceil^{n-1} \left(\lceil 0.5(\tau^I)^2 {L_I}^{-2} \rceil + 2n \lceil 0.25(\tau^I)^2 {L_I}^{-2} \rceil \right)$

Second-Order Clustering in X_3

Lemma: Suppose x^* is a constrained minimizer and $\exists \alpha > 0, \gamma_1 > 0, \gamma_2 > 0,$ and a set $\mathcal{D}_1$ s.t.

$$\nabla f(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 f(x^*) d \geq \gamma_1 d^{\mathrm{T}} d, \quad \forall d \in \mathcal{D}_1 \cap \mathcal{D}_I,$$

$$\max\left\{ \max_{j \in \mathcal{A}(x^*)} \left\{ \nabla g_j(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 \mathrm{g}_j(x^*) d \right\}, \max_{k \in \{1,\cdots,m_E\}} \left\{ \left| \nabla h_k(x^*)^{\mathrm{T}} d + \frac{1}{2} d^{\mathrm{T}} \nabla^2 \mathrm{h}_k(x^*) d \right| \right\} \right\} \geq \gamma_2 d^{\mathrm{T}} d, \quad \forall d \in \mathcal{D}_I \setminus \mathcal{D}_1,$$

where $\mathcal{D}_I$ is defined as

$$\mathcal{D}_I = \left\{ d : (x^* + d) \in \mathcal{N}_\alpha^2(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) \right\}.$$

Then $\exists \hat{\alpha} \in (0, \alpha]$ s.t. the region

$$\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_3 \cap \left\{ x = (x^* + d) \in \mathcal{N}_{\hat{\alpha}}^2(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_1 \cap \mathcal{D}_I \right\}$$

is overestimated by

$$\hat{X}_3^1 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \gamma_1 \left\| x - x^* \right\|_2^2 \leq 2\varepsilon^{\mathrm{o}} \right\}$$

and the region

$$\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_3 \cap \left\{ x = (x^* + d) \in \mathcal{N}_{\hat{\alpha}}^2(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1 \right\}$$

is overestimated by

$$\hat{X}_3^2 = \left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \gamma_2 \left\| x - x^* \right\|_2^2 \leq 2\varepsilon^{\mathrm{f}} \right\}.$$

Furthermore, suppose x^* is at the center of a box, B_δ, of width $\delta = \left(\frac{\varepsilon}{\tau^*} \right)^{\frac{1}{\beta^*}}$ placed while covering $\hat{X}_5$.

Then the region

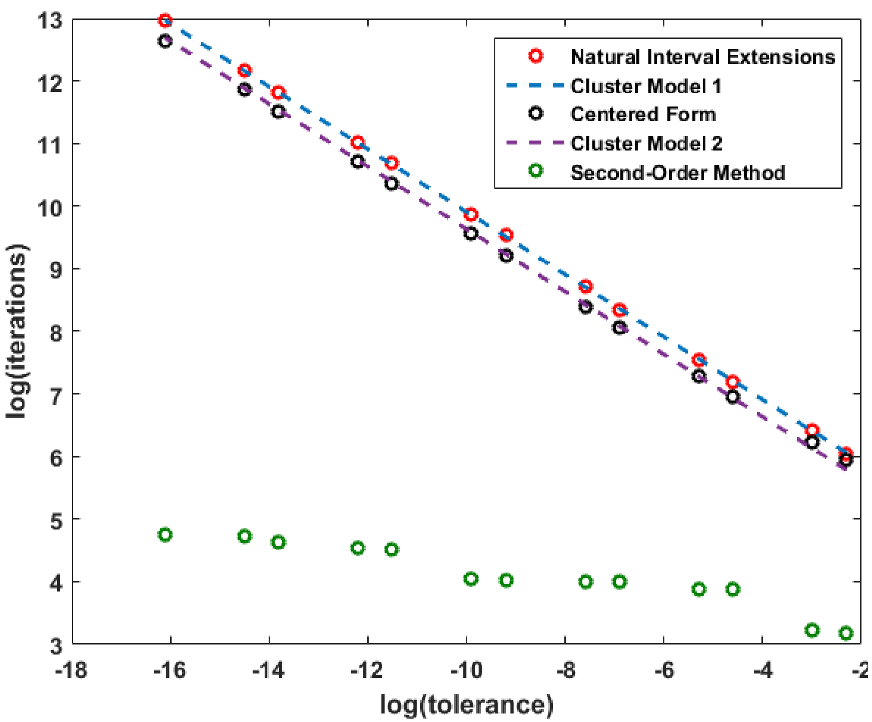
$$\mathcal{N}_{\hat{\alpha}}^2(x^*) \cap X_3 \cap \left\{ x = (x^* + d) \in \mathcal{N}_{\hat{\alpha}}^2(x^*) \cap \mathcal{F}^{\mathrm{C}}(X) : d \in \mathcal{D}_I \setminus \mathcal{D}_1 \right\} \setminus B_\delta$$

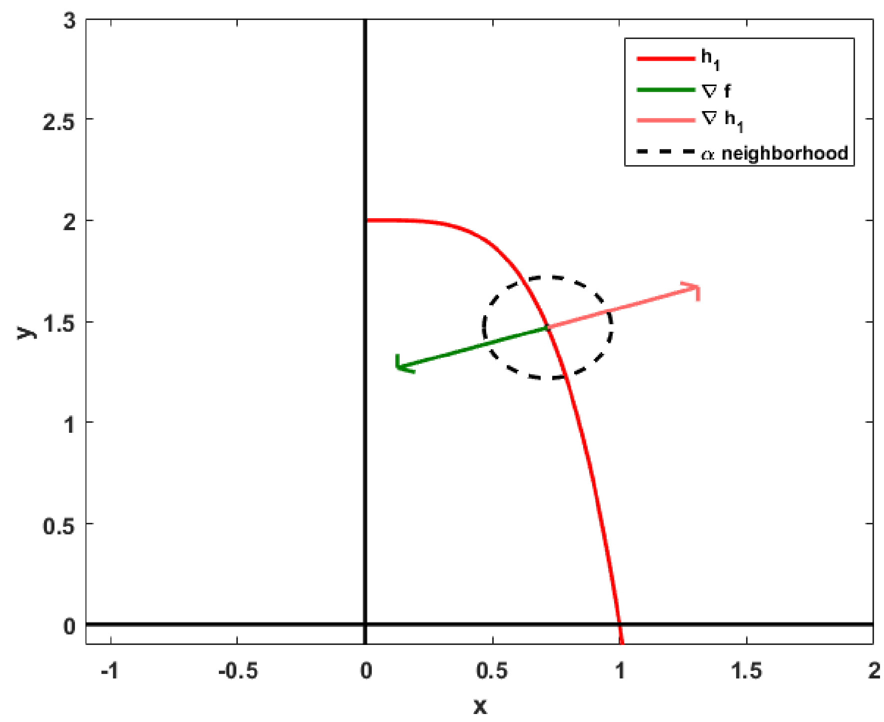
is overestimated by

$$\left\{ x \in \mathcal{N}_{\hat{\alpha}}^2(x^*) : \max\left\{ d\left(\{g(x)\}, \mathbb{R}_-^{m_I} \right), d\left(\{h(x)\}, \{0\} \right) \right\} \in \left[\frac{\gamma_2}{8} \delta^2, \varepsilon^{\mathrm{f}} \right] \right\}.$$

Revisiting the motivating examples

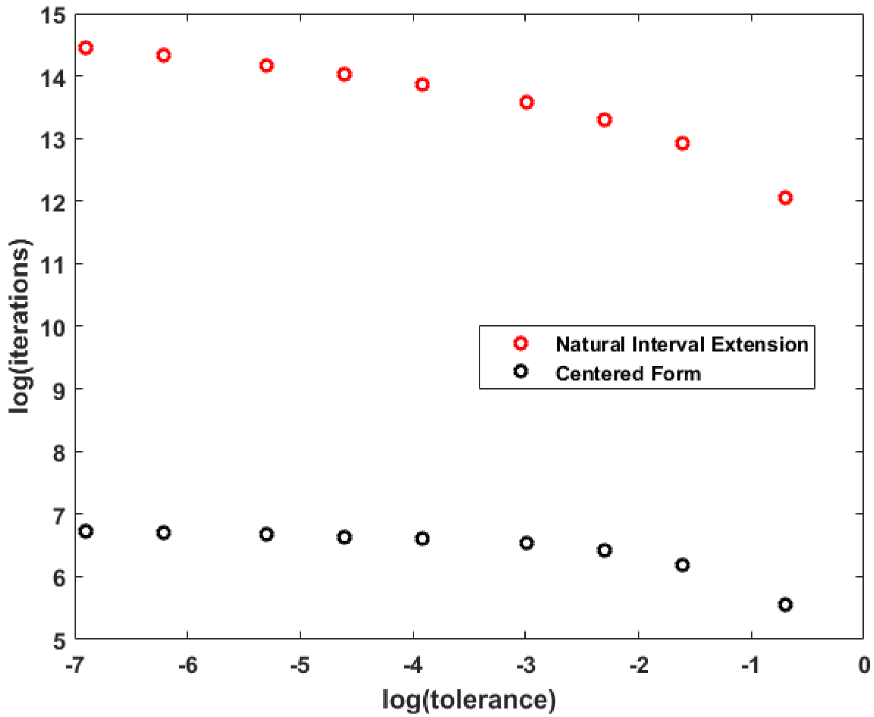
$$\min_{x,y}\ y^2 - 12x - 7y$$
$$\text{s.t.}\ \ y + 2x^4 - 2 = 0,$$
$$x \in [0,2],\ y \in [0,3].$$

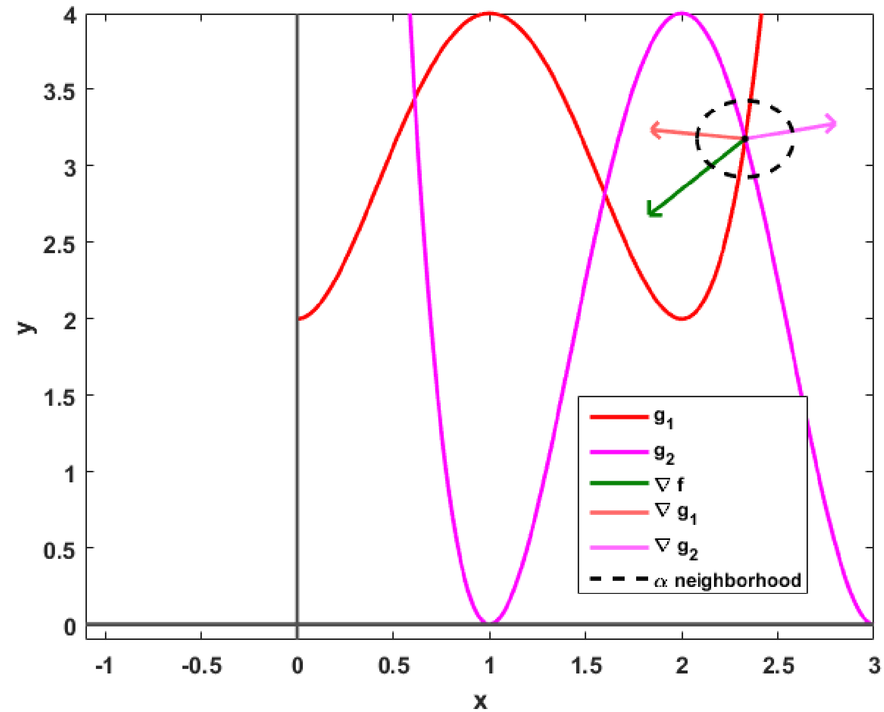




Revisiting the motivating examples

$$\min_{x,y} \; -x-y$$
$$\text{s.t. } y \le 2+2x^4-8x^3+8x^2,$$
$$y \le 4x^4-32x^3+88x^2-96x+36,$$
$$x\in[0,3],\; y\in[0,4].$$





Summary

- Illustrated the cluster problem (or lack thereof) in constrained optimization as motivation for analysis
- Proposed a notion of convergence order for convex relaxation-based lower bounding schemes for constrained problems
- Established sufficient conditions for first-order and second-order convergence of convex relaxation-based lower bounding schemes to mitigate clustering

Acknowledgements

- The Barton lab