Derivative-Free Robust Optimization by Outer Approximations

Stefan Wild
Mathematics and Computer Science Division
Argonne National Laboratory

Joint work with *Goldfarb grandson* **Matt Menickelly** (Argonne)

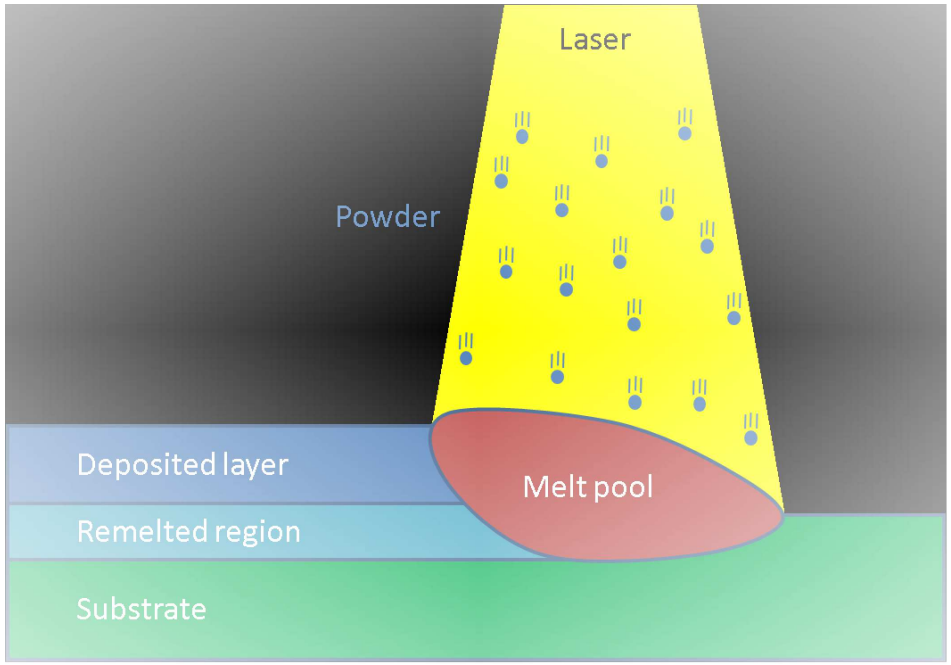
+ Sven Leyffer, Todd Munson, Charlie Vanaret (Argonne)

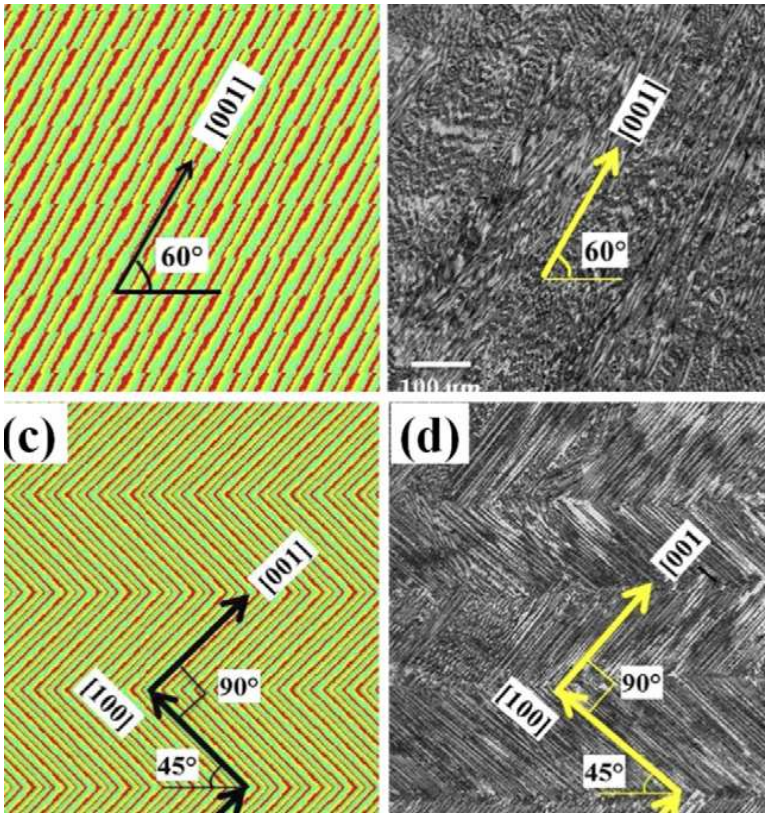
January 11, 2018

Outline

- Nonlinear robust optimization
- E. Polak's method of inexact outer approximation
- ∇f-free outer approximation
- Early numerical experience

$$\min_{\mathbf{x}\in\mathbb{R}^{\mathbf{n}}} \max_{\mathbf{u}\in\mathcal{U}} \mathbf{f}(\mathbf{x},\mathbf{u})$$





Images: [DebRoy, Zhang, Turner, Babu; ScrMat, 2017]

Nonlinear Robust Optimization

Guard against worst-case uncertainty in the problem data

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) \ : \ c(x, u) \leq 0 \qquad \forall u \in \mathcal{U} \right\}$$

where

f	certain objective
$c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$	uncertain constraints
u	uncertain variables/data
$\mathcal{U} \subset \mathbb{R}^m$	uncertainty set (compact, convex)

Well studied for linear (convex/concave) f, c

[Ben-Tal, El Ghaoui, Nemirovski; 2009], [Bertsimas, Brown, Caramani; SIRev 2011], ...

Nonlinear Robust Optimization

Guard against worst-case uncertainty in the problem data

$$\min_{x \in \mathbb{R}^n} \left\{ f(x) \ : \ c(x, u) \leq 0 \qquad \forall u \in \mathcal{U} \right\}$$

where

f certain objective

$c : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$ uncertain constraints

u uncertain variables/data

$\mathcal{U} \subset \mathbb{R}^m$ uncertainty set (compact, convex)

Well studied for linear (convex/concave) f, c

[Ben-Tal, El Ghaoui, Nemirovski; 2009], [Bertsimas, Brown, Caramani; SIRev 2011], ...

Special cases:

Minimax

$$\min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} f(x, u)$$

Implementation errors

$$\min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} f(x + u)$$

Another Case: Goldfarb Robust Optimization

Robust convex quadratically constrained programs

$$\min_{x \in \mathbb{R}^n} \left\{ c^\top x \; : \; \frac{1}{2} x^\top Q x + x^\top g + \gamma \leq 0 \quad \forall (Q, g, \gamma) \in \mathcal{U} \right\} \qquad \text{(RCQP)}$$

Another Case: Goldfarb Robust Optimization

Robust convex quadratically constrained programs

$$\min_{x\in\mathbb{R}^n}\left\{c^\top x \;:\; \frac{1}{2}x^\top Qx + x^\top g + \gamma \le 0 \quad \forall (Q,g,\gamma)\in\mathcal{U}\right\} \qquad \text{(RCQP)}$$

- [Ben-Tal, Nemirovski; MathOR, 1997]: $\mathcal{U}^i$ conditions to obtain SDP for (RCQP)
- [Goldfarb, Iyengar; MathProg, 2003]: $\mathcal{U}^i$ conditions to obtain SOCP for (RCQP)
  - Discrete/polytopic uncertainty sets

$$\mathcal{U} = \left\{(Q,g,\gamma)\;:(Q,g,\gamma) = \sum_{i=1}^{p}\lambda_i(Q^i,g^i,\gamma^i), \lambda\in\mathbb{R}^p_+, Q^i\succeq 0\,\forall i\;, \lambda^\top e = 1\right\}$$

  - Affine uncertainty sets $\mathcal{U}$

$$Q = Q^0 + \sum_{i=1}^{p}\lambda_i Q^i, \quad \|\lambda\|\le 1, \quad Q^i\succeq 0\,\forall i$$

$$(g,\gamma) = (g^0,\gamma^0) + \sum_{i=1}^{p} v_i(g^i,\gamma^i), \quad \|v\|\le 1$$

  - Factorized uncertainty sets $\mathcal{U}$

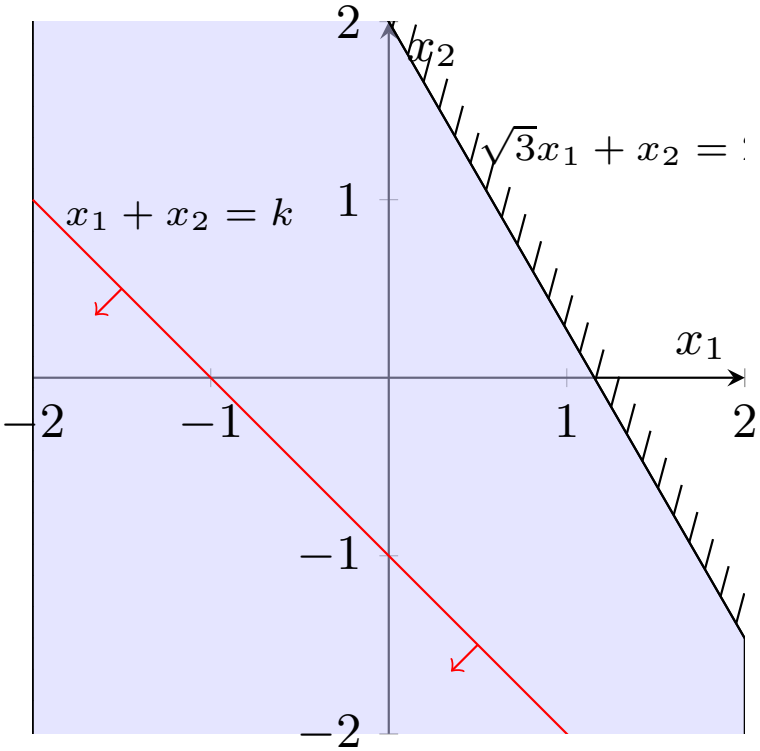
...

...CRs around MLEs

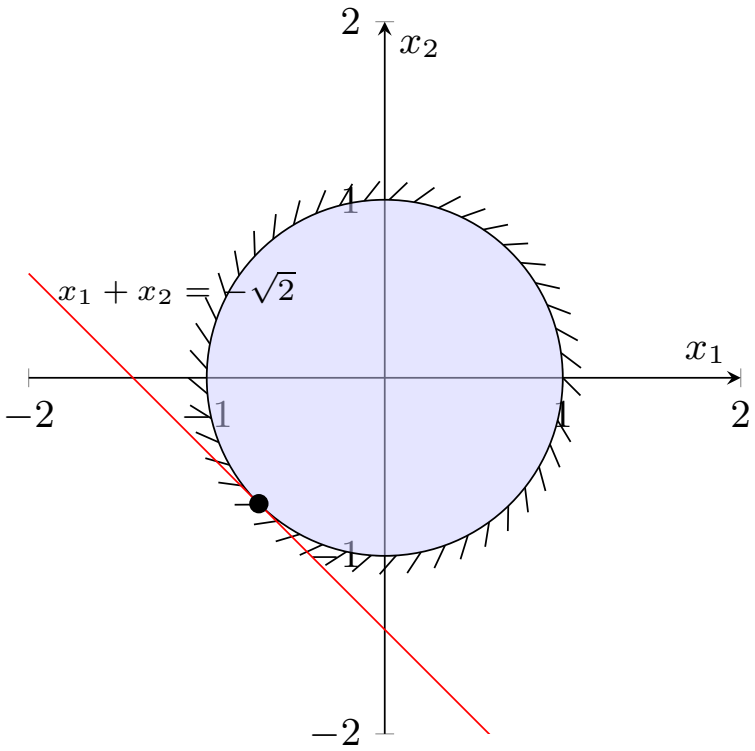
- See also *Robust portfolio selection problems* [Goldfarb, Iyengar; MOR, 2003]

Example of Robustness "Helping"

$$\min_{x\in\mathbb{R}^2}\left\{x_1+x_2 \;:\quad u_1x_1+u_2x_2-u_1^2-u_2^2\le 0,\ \forall u\in\mathcal{U}=[-1,1]^2\right\}$$



Nominal problem, $\hat{u}=(\frac{\sqrt{3}}{2},\frac{1}{2})$



Robust problem, $x^*=(-\frac{1}{\sqrt{2}},-\frac{1}{\sqrt{2}})$

Notation and Assumptions

Implicitly robustified form:

$$\min_{x \in \mathbb{R}^n} \max_{u \in \mathcal{U}} f(x, u) =: \min_{x \in \mathbb{R}^n} \Psi_{\mathcal{U}}(x) \tag{MM}$$

where, for any subset $\hat{\mathcal{U}} \subseteq \mathcal{U}$ use the relaxation:

$$\Psi_{\hat{\mathcal{U}}}(x) := \max_{u \in \hat{\mathcal{U}}} f(x, u) \qquad \leq \Psi_{\mathcal{U}}(x)$$

Sometimes forget and write $\Psi := \Psi_{\mathcal{U}}$

Notation and Assumptions

Implicitly robustified form:

$$\min_{x\in\mathbb{R}^n} \max_{u\in\mathcal{U}} f(x,u) =: \min_{x\in\mathbb{R}^n} \Psi_{\mathcal{U}}(x) \tag{MM}$$

where, for any subset $\hat{\mathcal{U}} \subseteq \mathcal{U}$ use the relaxation:

$$\Psi_{\hat{\mathcal{U}}}(x) := \max_{u\in\hat{\mathcal{U}}} f(x,u) \qquad \leq \Psi_{\mathcal{U}}(x)$$

Sometimes forget and write $\Psi := \Psi_{\mathcal{U}}$

Assume the following about (MM):

a. **Local Lipschitz continuity of f and $\nabla_x f$ everywhere**
$f(\cdot,\cdot)$ and, for any $u \in \mathcal{U}$, partial gradient $\nabla_x f(\cdot, u)$ Lipschitz continuous over any bounded subset of $\mathbb{R}^n \times \mathbb{R}^m$ and $\mathbb{R}^n$, resp.

b. **Compactness of $\mathcal{U}$**

c. (MM) **solution exists**

$\rightarrow$ no convexity of f or $\mathcal{U}$ assumed

An Optimality Measure

Employ second-order convex approximation of $f(\cdot, u)$ at x:

$$\Theta(x) \quad := \min_{h \in \mathbb{R}^n} \max_{u \in \mathcal{U}} \left\{ f(x,u) + \langle \nabla_x f(x,u), h \rangle + \frac{1}{2}\|h\|^2 \right\} - \Psi(x)$$

An Optimality Measure

Employ second-order convex approximation of $f(\cdot, u)$ at x:

$$\Theta(x) \quad := \min_{h \in \mathbb{R}^n} \max_{u \in \mathcal{U}} \left\{ f(x,u) + \langle \nabla_x f(x,u), h \rangle + \frac{1}{2}\|h\|^2 \right\} - \Psi(x)$$

Properties of Θ

For all $x \in \mathbb{R}^n$

1. $\Theta(x) \leq 0$
2. $\Theta(x)$ is continuous
3. $\mathbf{0} \in \partial\Psi(x)$ if and only if $\Theta(x) = 0$
4. $\Theta(x) =$
$$-\min_{\xi_0, \xi} \left\{ \xi_0 + \frac{1}{2}\|\xi\|^2 : \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} \in \mathbf{co}\left(\begin{bmatrix} \Psi_{\mathcal{U}}(x) - f(x,u) \\ \nabla_x f(x,u) \end{bmatrix} : u \in \mathcal{U} \right) \right\}$$

An Optimality Measure

Employ second-order convex approximation of $f(\cdot, u)$ at x:

$$\Theta(x) \quad := \min_{h \in \mathbb{R}^n} \max_{u \in \mathcal{U}} \left\{ f(x,u) + \langle \nabla_x f(x,u), h \rangle + \frac{1}{2}\|h\|^2 \right\} - \Psi(x)$$

Properties of Θ

For all $x \in \mathbb{R}^n$

1. $\Theta(x) \leq 0$
2. $\Theta(x)$ is continuous
3. $\mathbf{0} \in \partial\Psi(x)$ if and only if $\Theta(x) = 0$
4. $\Theta(x) =$
$$-\min_{\xi_0, \xi} \left\{ \xi_0 + \frac{1}{2}\|\xi\|^2 : \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} \in \mathbf{co}\left(\begin{bmatrix} \Psi_{\mathcal{U}}(x) - f(x,u) \\ \nabla_x f(x,u) \end{bmatrix} : u \in \mathcal{U} \right) \right\}$$

For any relaxation $\hat{\mathcal{U}} \subseteq \mathcal{U}$, will use

$$\Theta_{\hat{\mathcal{U}}}(x) := -\min_{\xi_0, \xi} \left\{ \xi_0 + \frac{1}{2}\|\xi\|^2 : \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} \in \mathbf{co}\left(\begin{bmatrix} \Psi_{\hat{\mathcal{U}}}(x) - f(x,u) \\ \nabla_x f(x,u) \end{bmatrix} : u \in \hat{\mathcal{U}} \right) \right\}$$
$$\leq \Theta(x) = \Theta_{\mathcal{U}}(x)$$

Inexact Method of Outer Approximation

Cutting-plane method from [Polak Optimization; 1997]

Uses approximate solutions of alternating block subproblems

$$\left(\min_{x \in \mathbb{R}^n} \Psi_{\hat{\mathcal{U}}}(x), \qquad \max_{u \in \mathcal{U}} f(\hat{x}, u) \right)$$

IOA Alg: Given data $\left\{ \left(\epsilon_k, \Omega^k \right) \right\}_{k=0}^{\infty}$

Initialize $x^0 \in \mathbb{R}^n$, $u^1 \in \operatorname*{argmax}_{u \in \Omega^0} f(x^0, u)$, $\mathfrak{U}^0 \leftarrow \{u^1\}$

Loop over k:

1. Compute any x^{k+1} such that $\Theta_{\mathfrak{U}^k}(x^{k+1}) \geq -\epsilon_k$
2. Compute any $u' \in \operatorname*{argmax}_{u \in \Omega^k} f(x^{k+1}, u)$ exactly
3. Augment $\mathfrak{U}^{k+1} \leftarrow \mathfrak{U}^k \cup \{u'\}$

Assumes:

a. $\Omega^k \subseteq \mathcal{U}$ and $\epsilon_k \in [0, 1]$ with $\lim_{k \to \infty} \epsilon_k = 0$

b. Ω^k grows dense in $\mathcal{U}$

c. $\min_{x \in \mathbb{R}^n} \max_{u \in \Omega^k} f(x, u)$ has a solution for all k

Result

Theorem [Polak]

Given assumptions on f and IOA Alg. Then, for any accumulation point x^* of $\{x^k\}_{k=1}^{\infty}$, $\Theta(x^*) = 0$. Thus, $\mathbf{0} \in \partial\Psi(x^*)$.

Basic idea is that as IOA progresses:

1. sequence of finite max functions

$$\Psi_{\Omega^k}(x) = \max_{u \in \Omega^k} f(x, u)$$

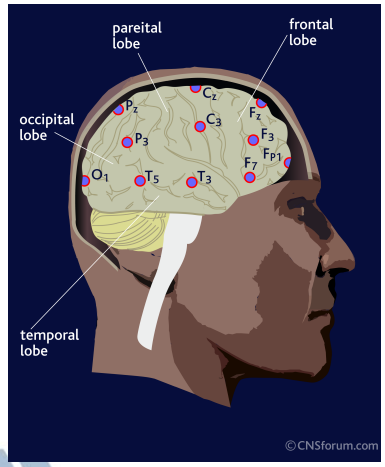
   are arbitrarily good approximations of $\Psi(x)$
2. sequence of optimality measures $\Theta_{\Omega^k}(x)$ are arbitrarily good approximations of the optimality measure $\Theta(x)$

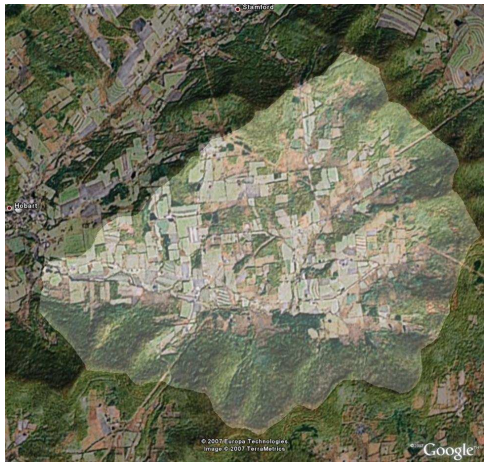
When the Derivatives Start Hiding: Simulation-Based Optimization

$$\min_{x \in \mathbb{R}^n} \{h(x; S(x)) : c_I[x, S(x)] \leq 0,\ c_E[x, S(x)] = 0\}$$

- $S : \mathbb{R}^n \to \mathbb{C}^p$ simulation output, often "noisy" (even when deterministic)
- Derivatives $\nabla_x S$ often unavailable or prohibitively expensive to obtain/approximate directly
- S can contribute to objective and/or constraints
- Single evaluation of S could take seconds/minutes/hours/...
  $\Rightarrow$ Evaluation is a bottleneck for optimization
- This talk: $h(x; S(x)) = \max_{u \in \mathcal{U}} f(x, u)$

Functions of complex (numerical) simulations arise everywhere





Derivative-Free Inexact Outer Approximation

Main task:

Compute sufficiently accurate approximation of

$$\Theta_{\Omega^k}(x^k) = -\min_{\xi_0,\xi}\left\{\xi_0 + \frac{\|\xi\|^2}{2} : \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} \in \mathbf{co}\left(\begin{bmatrix} \Psi_{\Omega^k}(x^k) - f(x^k,u) \\ \nabla_x f(x^k,u) \end{bmatrix} : u \in \Omega^k\right)\right\}$$

for which $\Theta_{\Omega^k}(x^k) \leq \epsilon_k$ is attainable when

- ∇f values unavailable
- $f(x,u)$ evaluations expensive

Derivative-Free Inexact Outer Approximation

Main task:

Compute sufficiently accurate approximation of

$$\Theta_{\Omega^k}(x^k) = -\min_{\xi_0,\xi}\left\{\xi_0 + \frac{\|\xi\|^2}{2} : \begin{bmatrix} \xi_0 \\ \xi \end{bmatrix} \in \mathbf{co}\left(\begin{bmatrix} \Psi_{\Omega^k}(x^k) - f(x^k,u) \\ \nabla_x f(x^k,u) \end{bmatrix} : u \in \Omega^k\right)\right\}$$

for which $\Theta_{\Omega^k}(x^k) \leq \epsilon_k$ is attainable when

- ∇f values unavailable
- $f(x,u)$ evaluations expensive

Approach

Phase 1 Inner iterations to obtain x^{k+1} an approximate minimizer of

$$\min_x \Psi_{\mathfrak{U}^k}(x)$$

$\rightarrow$ Manifold sampling, trust-region approach

Phase 2 Solve $\operatorname{argmax}_{u\in\Omega^k} f(x^{k+1},u)$

Model-Based Approximation for Inner Solve of $\min_x \Psi_{\mathfrak{U}^k}(x)$

Associate with each $u^j \in \mathfrak{U}^k$ a model about primal iterate y^t ($y^t \to^t x^{k+1}$):

Fully Linear Models

m_j^t fully linear model of $f(\cdot, u^j)$ on $\mathcal{B}(y^t, \Delta)$ if there exist constants $\kappa_{j,\mathrm{ef}}$ and $\kappa_{j,\mathrm{eg}}$ independent of y^t and Δ with

$$|f(y^t + s, u^j) - m_j^t(y^t + s)| \leq \kappa_{j,\mathrm{ef}} \Delta^2 \quad \forall s \in \mathcal{B}(0, \Delta)$$
$$\|\nabla_x f(y^t + s, u^j) - \nabla m_j^t(y^t + s)\| \leq \kappa_{j,\mathrm{eg}} \Delta \quad \forall s \in \mathcal{B}(0, \Delta)$$

[Conn, Scheinberg, Vicente; SIAM, 2009]

Model-Based Approximation for Inner Solve of $\min_x \Psi_{\mathfrak{U}^k}(x)$

Associate with each $u^j \in \mathfrak{U}^k$ a model about primal iterate y^t ($y^t \to^t x^{k+1}$):

Fully Linear Models

m_j^t fully linear model of $f(\cdot, u^j)$ on $\mathcal{B}(y^t, \Delta)$ if there exist constants $\kappa_{j,\mathrm{ef}}$ and $\kappa_{j,\mathrm{eg}}$ independent of y^t and Δ with

$$\begin{aligned} |f(y^t + s, u^j) - m_j^t(y^t + s)| &\le \kappa_{j,\mathrm{ef}} \Delta^2 \quad \forall s \in \mathcal{B}(0, \Delta) \\ \|\nabla_x f(y^t + s, u^j) - \nabla m_j^t(y^t + s)\| &\le \kappa_{j,\mathrm{eg}} \Delta \quad \forall s \in \mathcal{B}(0, \Delta) \end{aligned}$$

[Conn, Scheinberg, Vicente; SIAM, 2009]

For set of *generator indices*, $J^{t,k} \subseteq \mathfrak{U}^k$:

- $f(y^t, u^j) = m_j^t(y^t)$ for all $j \in J^{t,k}$
- Fully linear

Both trivial (e.g., linear models) when ∇f is Lipschitz

$$\begin{aligned} G^t &:= \left[\nabla m_{\sigma(1)}^t(y^t), \ldots, \nabla m_{\sigma(|J^{t,k}|)}^t(y^t)\right] \in \mathbb{R}^{n \times |J^{t,k}|} \\ F^t &:= \left[f(y^t, u^{\sigma(1)}), \ldots, f(y^t, u^{\sigma(|J^{t,k}|)})\right]^\top \in \mathbb{R}^{|J^{t,k}|} \end{aligned}$$

Natural Approx

Use model-based set

$$\mathcal{D}_{m^t,\mathfrak{U}^k}(y^t) := \mathbf{co}\left\{ \begin{bmatrix} \Psi_{\mathfrak{U}^k}(y^t) - m_j^t(y^t) \\ \nabla_x m_j^t(y^t) \end{bmatrix} : u^j \in \mathfrak{U}^k \right\}$$

to define approximate inexact measure

$$\tilde{\Theta}^t_{\mathfrak{U}^k}(y^t) := -\min_{z_0,z} \left\{ z_0 + \frac{1}{2}\|z\|^2 : \begin{bmatrix} z_0 \\ z \end{bmatrix} \in \mathcal{D}_{m^t,\mathfrak{U}^k}(y^t) \right\}$$

Natural Approx

Use model-based set

$$\mathcal{D}_{m^t,\mathfrak{U}^k}(y^t) := \mathbf{co}\left\{\begin{bmatrix} \Psi_{\mathfrak{U}^k}(y^t) - m_j^t(y^t) \\ \nabla_x m_j^t(y^t) \end{bmatrix} : u^j \in \mathfrak{U}^k\right\}$$

to define approximate inexact measure

$$\tilde{\Theta}^t_{\mathfrak{U}^k}(y^t) := -\min_{z_0,z}\left\{z_0 + \frac{1}{2}\|z\|^2 : \begin{bmatrix} z_0 \\ z \end{bmatrix} \in \mathcal{D}_{m^t,\mathfrak{U}^k}(y^t)\right\}$$

Result- $\tilde{\Theta}^t_{\mathfrak{U}^k}(y^t)$ is a fully linear approximation of $\Theta_{\mathfrak{U}^k}(y^t)$.

For all $(z_0, z) \in \mathcal{D}_{m^t,\mathfrak{U}^k}(y^t)$, there exists $(\xi_0(z_0, z), \xi(z_0, z)) \in \mathcal{D}_{f,\mathcal{U}}(y^t)$ with

$$\begin{aligned} z_0 &= \xi_0(z_0, z) \\ \|z - \xi(z_0, z)\| &\leq \kappa_g \Delta_k. \end{aligned}$$

Note:

- $\mathcal{D}_{m^t,\mathfrak{U}^k}(y^t)$ relies on $|\mathfrak{U}^k|$
- In practice, ensure fully linear approximation of only $|J^{t,k}|$ many models in inner iteration t

Algorithm

DFOA Alg: Given data $\left\{\left(\epsilon_k, \Omega^k\right)\right\}_{k=0}^{\infty}$

Initialize $x^0 \in \mathbb{R}^n$, $u^1 \in \operatorname*{argmax}_{u \in \Omega^0} f(x^0, u)$, $\mathfrak{U}^0 \leftarrow \{u^1\}$

Loop over k:

- $t \leftarrow 0$; $y^t \leftarrow x^k$, $\Delta_t \leftarrow \Delta_{\text{init}}$; $\chi_t \leftarrow \infty$
- Inner iterations while $\chi_t > \epsilon_k$:
  1. **(Phase 1)** Choose set $J^{t,k}$ satisfying
     $$\operatorname*{argmax}_{j=1,\ldots,|\mathfrak{U}^k|} f(y^t, u^j) \subseteq J^{t,k} \subseteq \{1, \ldots, |\mathfrak{U}^k|\}$$
  2. Build models $\{m_j^t : j \in J^{t,k}\}$ and solve TRSP
     $$\min_{(z,d) \in \mathbb{R}^{1+n}} \left\{ z + \frac{1}{2} d^\top B^t d : F^t - \Psi_{\mathfrak{U}^k}(y^t)\mathbf{e} + (G^t)^\top d \leq z\mathbf{e}, \ \|d\| \leq \Delta_t \right\}$$
  3. If $J^*(y^t + d^t) \in J^{t,k}$ perform trust region update
- **(Phase 2)** Compute $u' \in \operatorname*{argmax}_{u \in \Omega^k} f(x^{k+1}, u)$

  Augment $\mathfrak{U}^{k+1} \leftarrow \mathfrak{U}^k \cup \{u'\}$

Details

Stopping criterion for inner iterations

Employ dual measure

$$
\begin{aligned}
\chi^t &= \min_{z_0, z} \left\{ z_0 + \frac{1}{2}\|z\| : \begin{bmatrix} z_0 \\ z \end{bmatrix} \in \mathcal{D}_{m^t, \mathfrak{U}^k_{J^{t,k}}}(y^t) \right\} \\
&\geq -\tilde{\Theta}^t_{\mathfrak{U}^k}(y^t) \\
&= \min_{z_0, z} \left\{ z_0 + \frac{1}{2}\|z\|^2 : \begin{bmatrix} z_0 \\ z \end{bmatrix} \in \mathcal{D}_{m^t, \mathfrak{U}^k}(y^t) \right\}
\end{aligned}
$$

Details

Stopping criterion for inner iterations

Employ dual measure

$$
\begin{aligned}
\chi^t &= \min_{z_0, z} \left\{ z_0 + \frac{1}{2}\|z\| : \begin{bmatrix} z_0 \\ z \end{bmatrix} \in \mathcal{D}_{m^t, \mathfrak{U}^k_{J^{t,k}}}(y^t) \right\} \\
&\geq -\tilde{\Theta}^t_{\mathfrak{U}^k}(y^t) \\
&= \min_{z_0, z} \left\{ z_0 + \frac{1}{2}\|z\|^2 : \begin{bmatrix} z_0 \\ z \end{bmatrix} \in \mathcal{D}_{m^t, \mathfrak{U}^k}(y^t) \right\}
\end{aligned}
$$

Step acceptance criterion

$$
\rho_t \triangleq \frac{\Psi_{\mathfrak{U}^k}(y^t) - \Psi_{\mathfrak{U}^k}(y^t + d^t)}{-(z^t + \frac{1}{2} d^{t\top} B^t d^t)}
$$

Main result for DFOA

Thm: Given assumptions on f and IOA Alg.
Then, for any accumulation point x^* of $\{x^k\}_{k=1}^{\infty}$, $\mathbf{0} \in \partial\Psi(x^*)$.

Idea

1. On acceptable inner iterations, if

$$\Delta_k < \min\left\{\min\left\{\frac{\kappa_{\mathrm{fcd}}(1-\eta_1)}{3\kappa_f + \frac{1}{2}\kappa_{\mathrm{mh}}}, \eta_2\right\}\chi_t, 1\right\}$$

   then inner iteration is successful
2. Δ_k tends to 0 in each inner iteration
3. For all $\epsilon_k > 0$, finite number of inner iterations to achieve $\chi_t < \epsilon_k$
4. For all $\epsilon_k > 0$, $\chi_t \leq \epsilon_k$ implies that

$$-\Theta_{\mathfrak{U}^k}(y^t) \leq \epsilon_k + \kappa_g \eta_2 \epsilon_k^2 + \frac{1}{2}\kappa_g^2 \eta_2^2 \epsilon_k^2$$

5. Appeal to Polak IOA

Important Practicalities

Selection of B^t (used in TRSP)

Φ_Q quadratic polynomial basis:

$$\Phi_Q(v) \triangleq \left[\frac{1}{2}v_1^2, \ldots, \frac{1}{2}v_n^2, v_1v_2, \ldots, v_2v_3, \ldots, v_{n-1}v_n\right]$$

obtain coeffs α_Q by least-squares soln

$$\begin{bmatrix} \Phi_Q(p^1) \\ \vdots \\ \Phi_Q(p^{|P|}) \end{bmatrix} \alpha_Q = \begin{bmatrix} \Psi_{\mathfrak{U}(J^{t,k})}(p^1) - \max\limits_{j=1,\ldots,|J^{t,k}|} \left\{F_j^t + (G_j^t)^\top (p^1 - y^t)\right\} \\ \vdots \\ \Psi_{\mathfrak{U}(J^{t,k})}(p^{|P|}) - \max\limits_{j=1,\ldots,|J^{t,k}|} \left\{F_j^t + (G_j^t)^\top (p^{|P|} - y^t)\right\} \end{bmatrix}$$

+ Does not require additional evaluations

Important Practicalities

Selection of B^t (used in TRSP)

Φ_Q quadratic polynomial basis:

$$\Phi_Q(v) \triangleq \left[\frac{1}{2}v_1^2, \ldots, \frac{1}{2}v_n^2, v_1v_2, \ldots, v_2v_3, \ldots, v_{n-1}v_n\right]$$

obtain coeffs α_Q by least-squares soln

$$\begin{bmatrix} \Phi_Q(p^1) \\ \vdots \\ \Phi_Q(p^{|P|}) \end{bmatrix} \alpha_Q = \begin{bmatrix} \Psi_{\mathfrak{U}(J^{t,k})}(p^1) - \max\limits_{j=1,\ldots,|J^{t,k}|} \left\{F_j^t + (G_j^t)^\top (p^1 - y^t)\right\} \\ \vdots \\ \Psi_{\mathfrak{U}(J^{t,k})}(p^{|P|}) - \max\limits_{j=1,\ldots,|J^{t,k}|} \left\{F_j^t + (G_j^t)^\top (p^{|P|} - y^t)\right\} \end{bmatrix}$$

+ Does not require additional evaluations

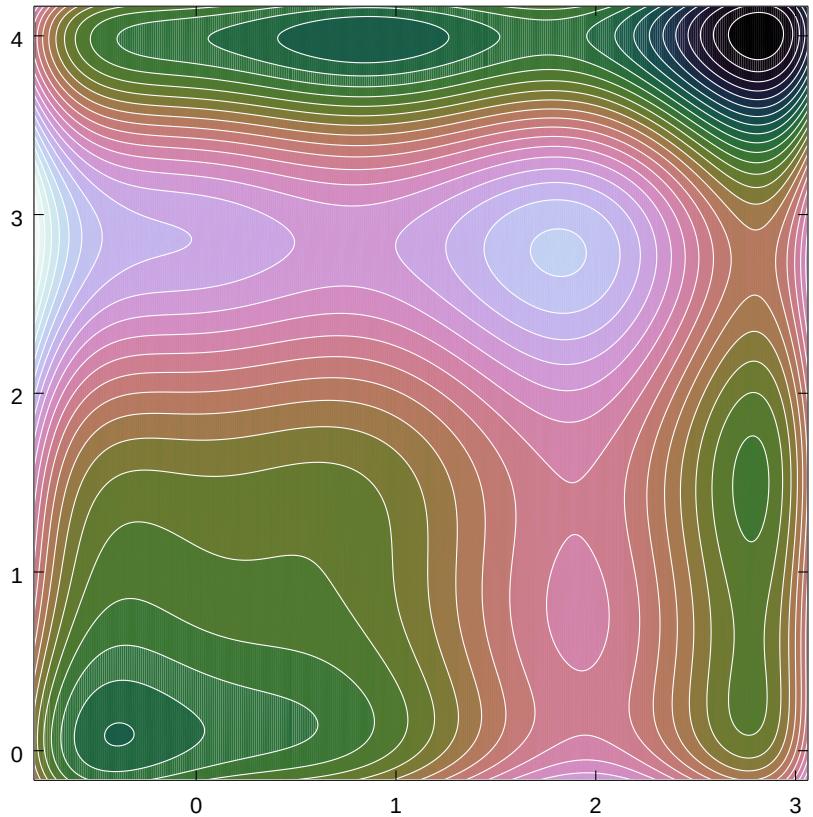
(Pre)Selection of $\{\Omega^k\}_{k=0}^{\infty}$

- Theory requires dense in $\mathcal{U}$ [Gonzaga, Polak; SICON, 1979]
- In experiments we consider very few points

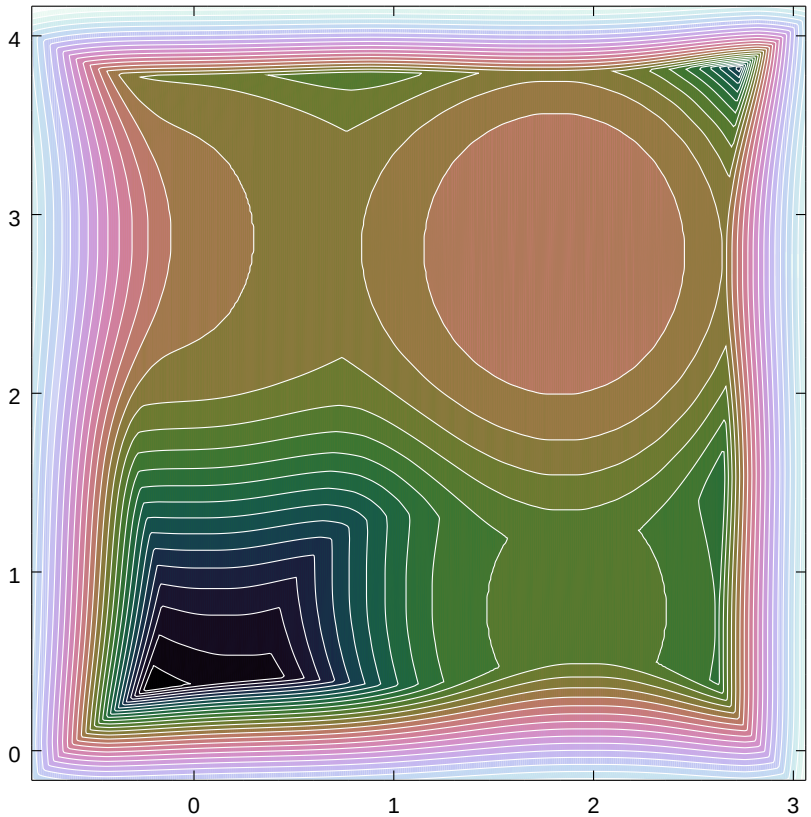
Bertsimas-Nohadani-Teo 2D Implementation Error Problem

$$\min_{x\in\mathbb{R}^2} \Psi_{\mathcal{U}_\alpha}(x) := \min_{x\in\mathbb{R}^2} \max_{u:\|u\|_2\le\alpha} f(x,u) := \min_{x\in\mathbb{R}^2} \max_{u:\|u\|_2\le\alpha} g(x+u)$$

parameter $\alpha \ge 0$ ($\alpha = 0.5$ being typical)



$g(x) = f(x,0)$



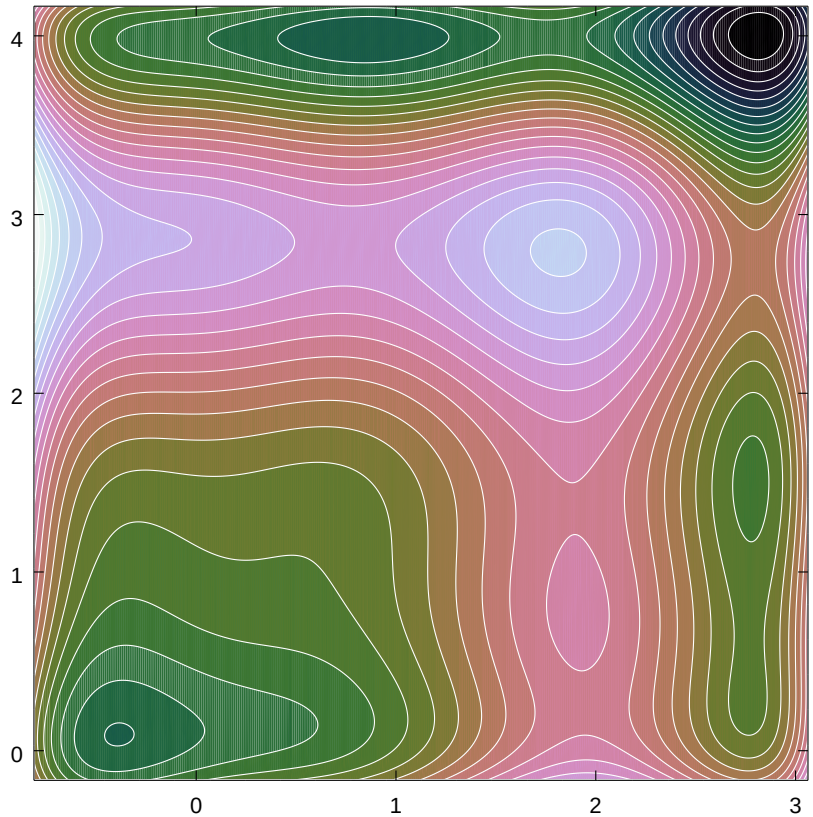
$\Psi_{\mathcal{U}_{0.5}}(x)$

$$\begin{aligned} g(x) \quad = \quad & 2x_1^6 - 12.2x_1^5 + 21.2x_1^4 - 6.4x_1^3 - 4.7x_1^2 + 6.2x_1 + x_2^6 - 11x_2^5 + 43.3x_2^4 \\ & -74.8x_2^3 + 56.9x_2^2 - 10x_2 - 0.1x_1^2 + x_2^2 + 0.4x_1^2x_2 + 0.4x_2^2x_1 - 4.1x_1x_2 \end{aligned}$$

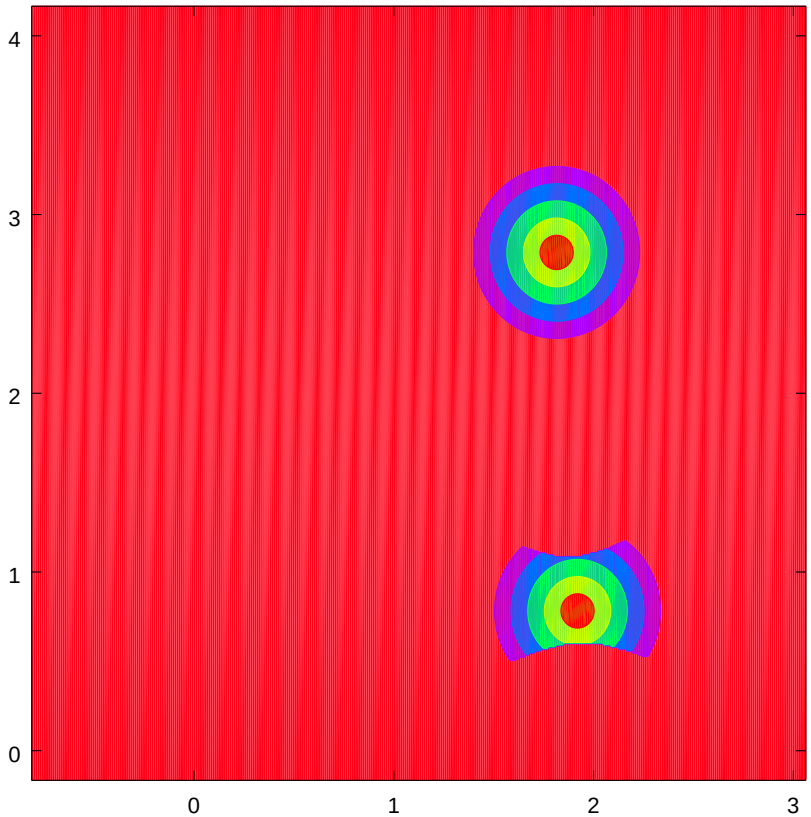
Bertsimas-Nohadani-Teo 2D Implementation Error Problem

$$\min_{x\in\mathbb{R}^2} \Psi_{\mathcal{U}_\alpha}(x) := \min_{x\in\mathbb{R}^2} \max_{u:\|u\|_2\le\alpha} f(x,u) := \min_{x\in\mathbb{R}^2} \max_{u:\|u\|_2\le\alpha} g(x+u)$$

parameter $\alpha \ge 0$ ($\alpha = 0.5$ being typical)



$g(x) = f(x,0)$



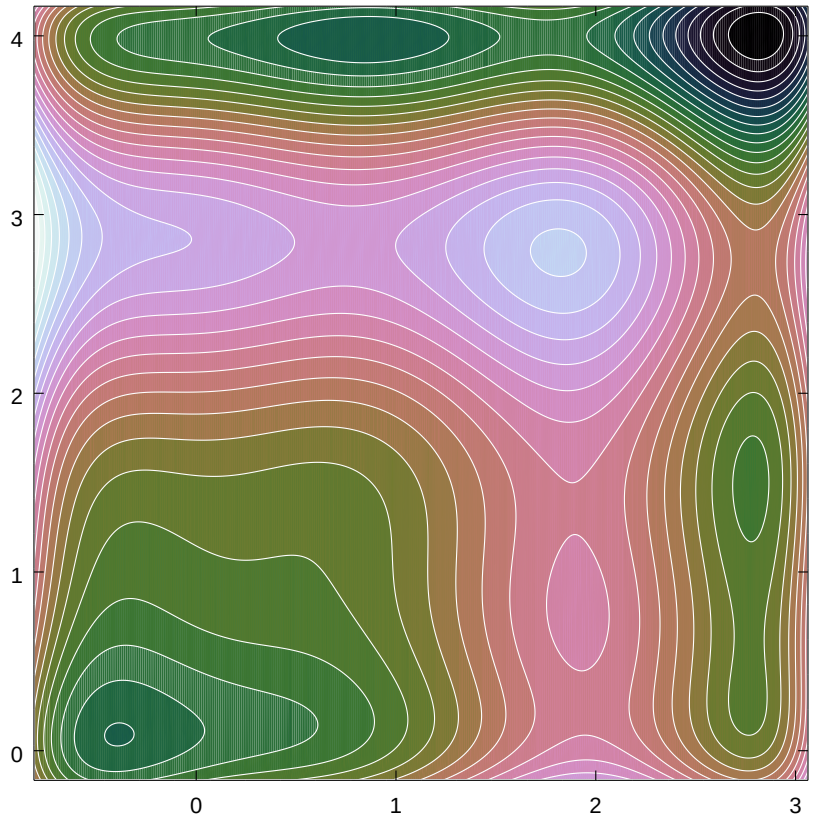
$\|u^*\|$

$$\begin{aligned} g(x) \quad = \quad & 2x_1^6 - 12.2x_1^5 + 21.2x_1^4 - 6.4x_1^3 - 4.7x_1^2 + 6.2x_1 + x_2^6 - 11x_2^5 + 43.3x_2^4 \\ & -74.8x_2^3 + 56.9x_2^2 - 10x_2 - 0.1x_1^2 + x_2^2 + 0.4x_1^2x_2 + 0.4x_2^2x_1 - 4.1x_1x_2 \end{aligned}$$

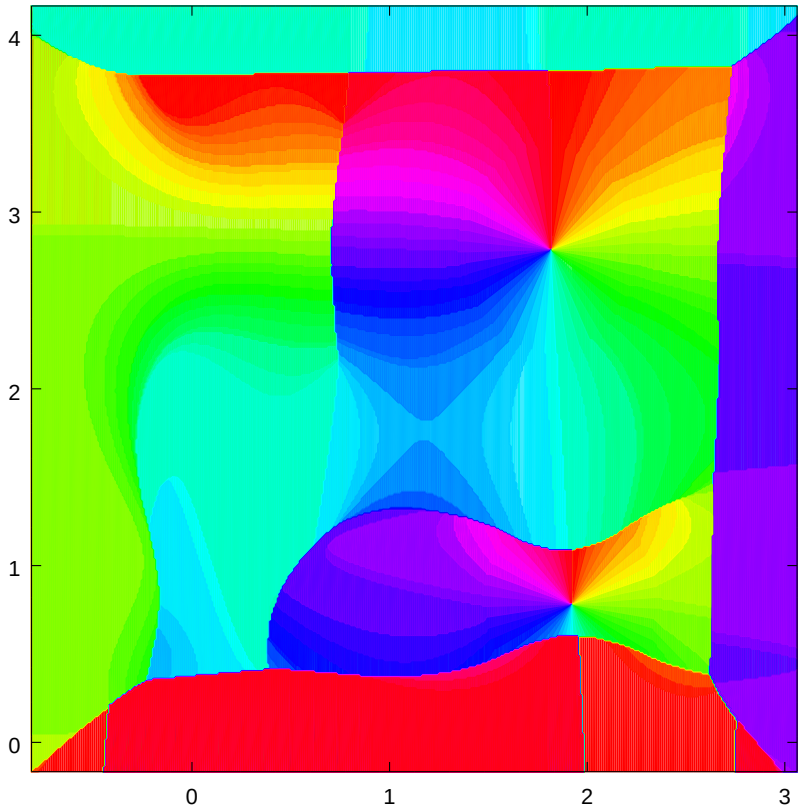
Bertsimas-Nohadani-Teo 2D Implementation Error Problem

$$\min_{x\in\mathbb{R}^2} \Psi_{\mathcal{U}_\alpha}(x) := \min_{x\in\mathbb{R}^2} \max_{u:\|u\|_2\le\alpha} f(x,u) := \min_{x\in\mathbb{R}^2} \max_{u:\|u\|_2\le\alpha} g(x+u)$$

parameter $\alpha \ge 0$ ($\alpha = 0.5$ being typical)



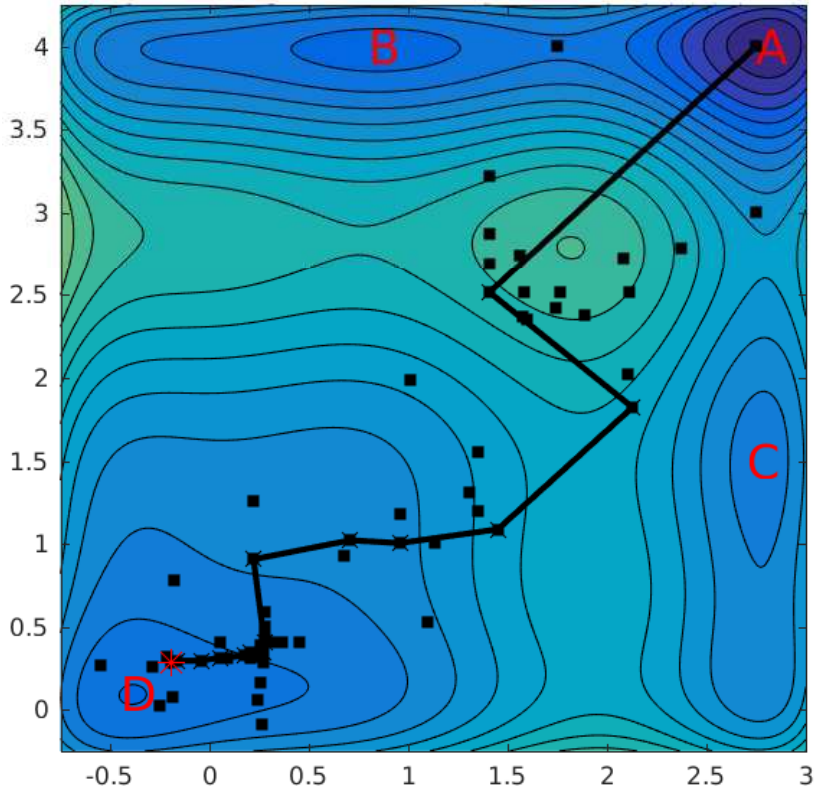
$g(x) = f(x,0)$



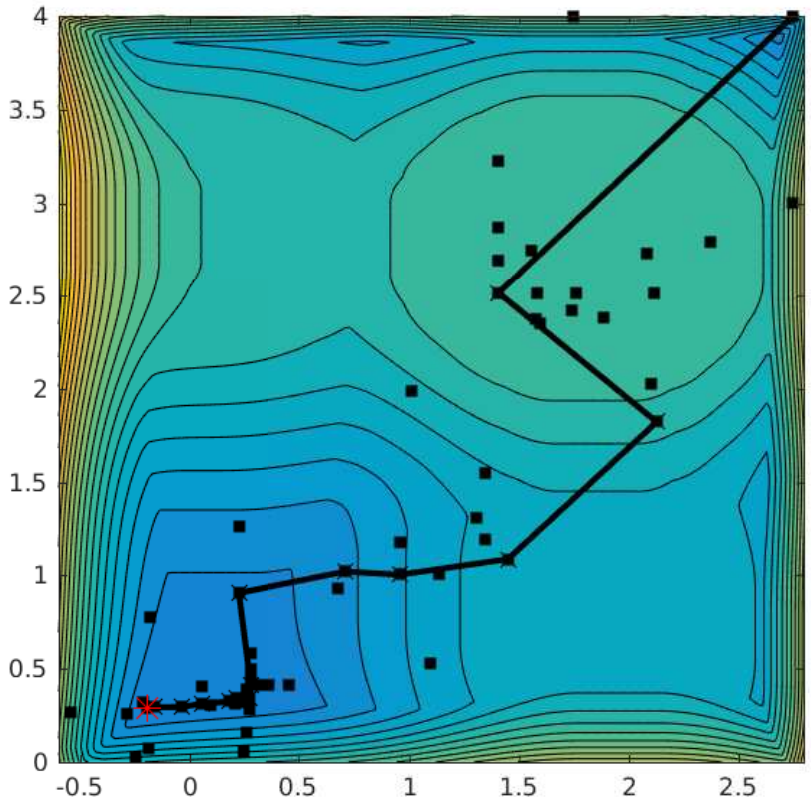
$\angle u^*$

$$\begin{aligned} g(x) \quad = \quad & 2x_1^6 - 12.2x_1^5 + 21.2x_1^4 - 6.4x_1^3 - 4.7x_1^2 + 6.2x_1 + x_2^6 - 11x_2^5 + 43.3x_2^4 \\ & -74.8x_2^3 + 56.9x_2^2 - 10x_2 - 0.1x_1^2 + x_2^2 + 0.4x_1^2x_2 + 0.4x_2^2x_1 - 4.1x_1x_2 \end{aligned}$$

DFOA Trajectories Bertsimas-Nohadani-Teo



$g(x) = f(x, 0)$

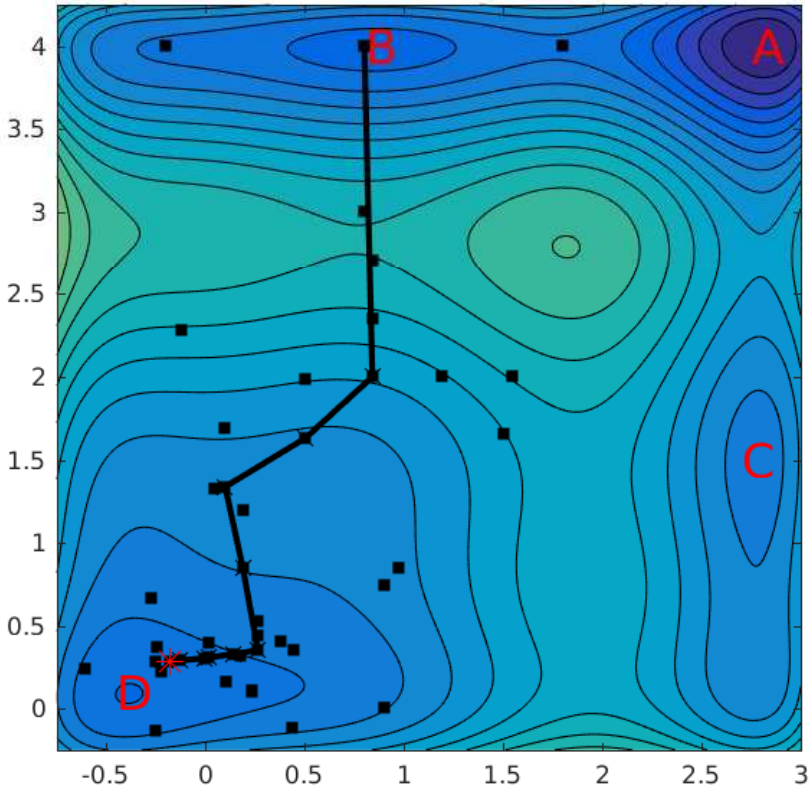


$\Psi_{\mathcal{U}_{0.5}}(x)$

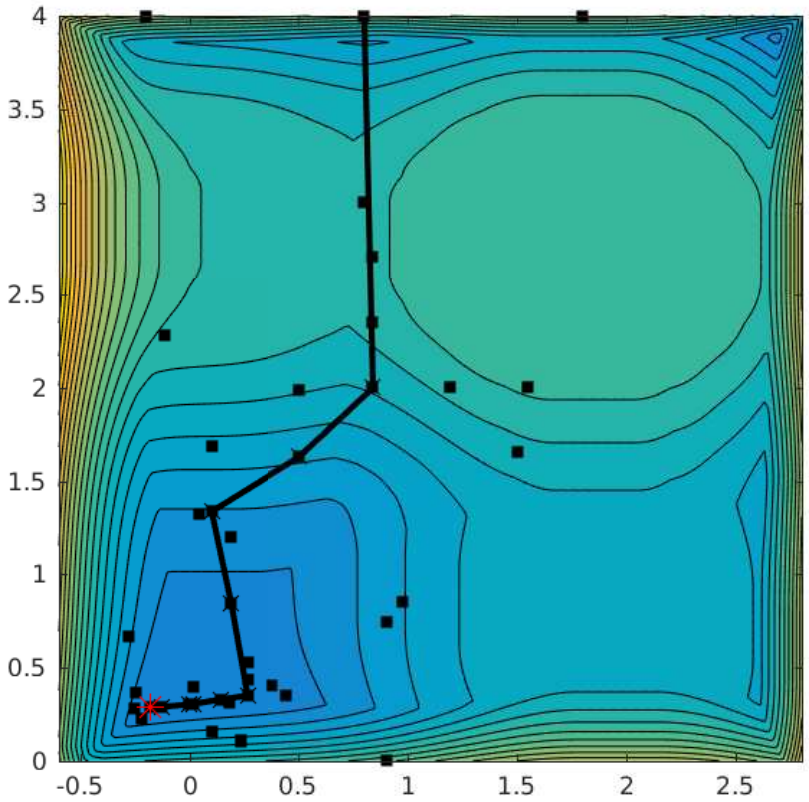
- Recover global within 250 total evaluations
- $\mathfrak{U}^0 = \{\pm 0.5 e_i : i = 1, 2\}$

See also [Bertsimas, Nohadani; JOGO 2010]; [Conn, Vicente; OMS 2012]

DFOA Trajectories Bertsimas-Nohadani-Teo



$g(x) = f(x, 0)$

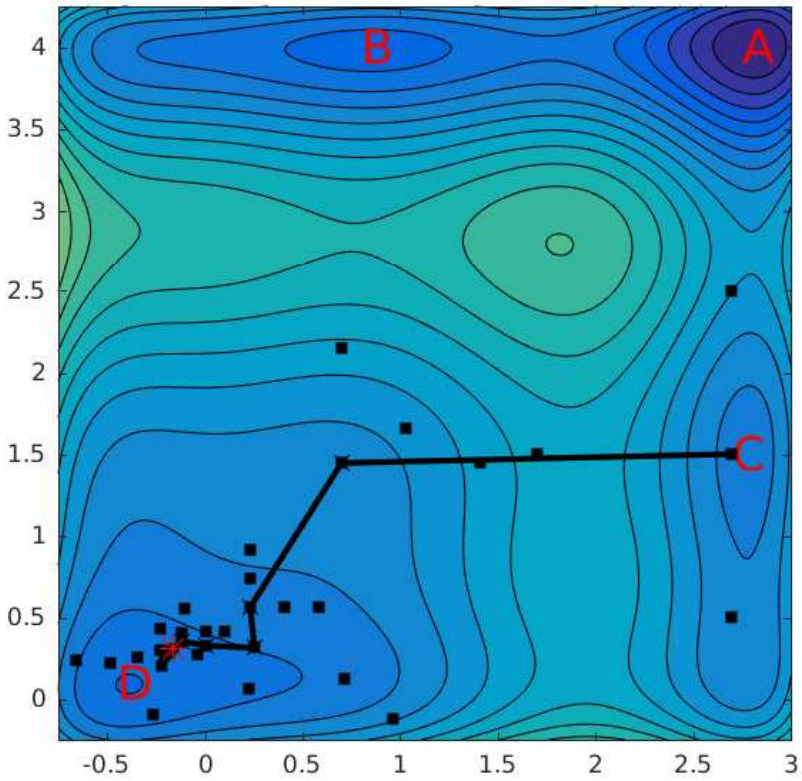


$\Psi_{\mathcal{U}_{0.5}}(x)$

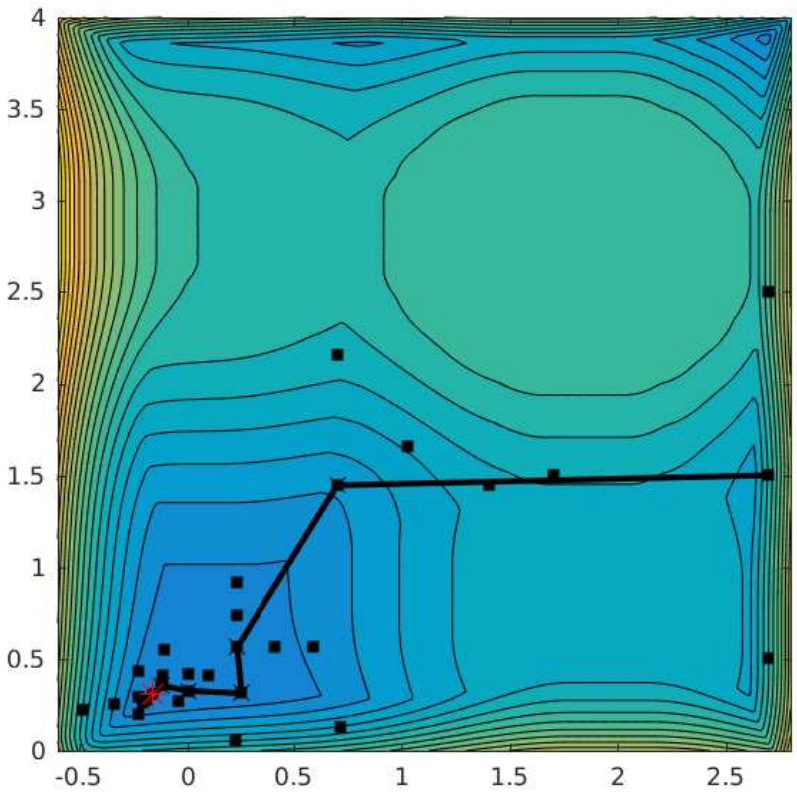
- Recover global within 250 total evaluations
- $\mathfrak{U}^0 = \{\pm 0.5 e_i : i = 1, 2\}$

See also [Bertsimas, Nohadani; JOGO 2010]; [Conn, Vicente; OMS 2012]

DFOA Trajectories Bertsimas-Nohadani-Teo



$g(x) = f(x, 0)$

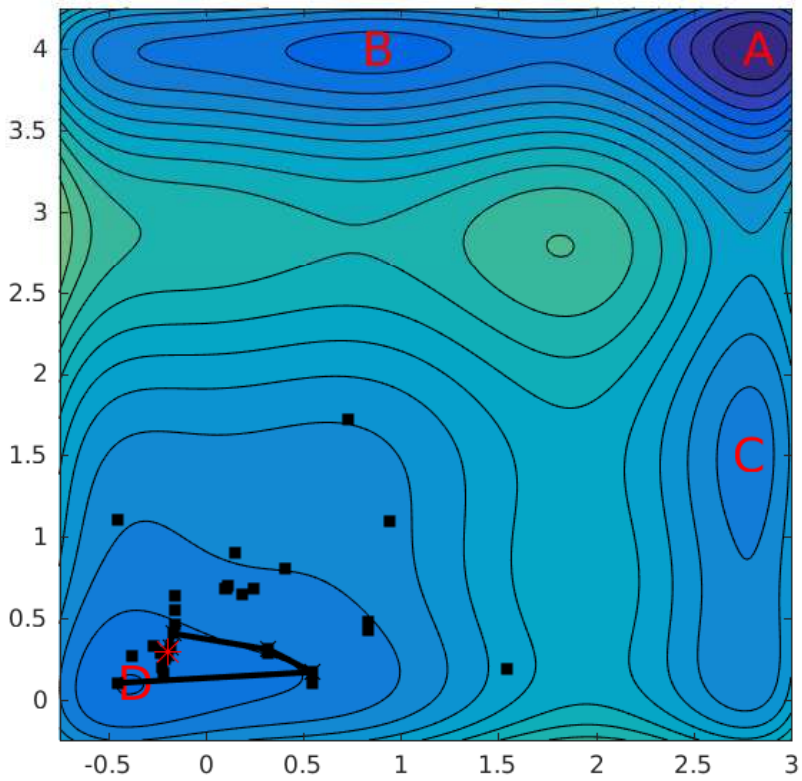


$\Psi_{\mathcal{U}_{0.5}}(x)$

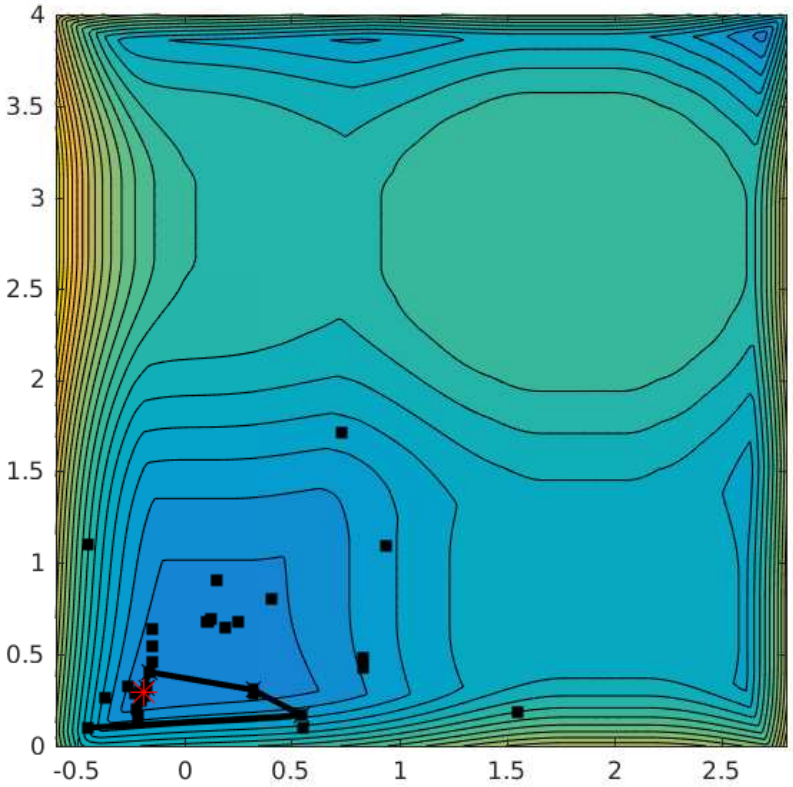
- Recover global within 250 total evaluations
- $\mathfrak{U}^0 = \{\pm 0.5 e_i : i = 1, 2\}$

See also [Bertsimas, Nohadani; JOGO 2010]; [Conn, Vicente; OMS 2012]

DFOA Trajectories Bertsimas-Nohadani-Teo



$g(x) = f(x, 0)$

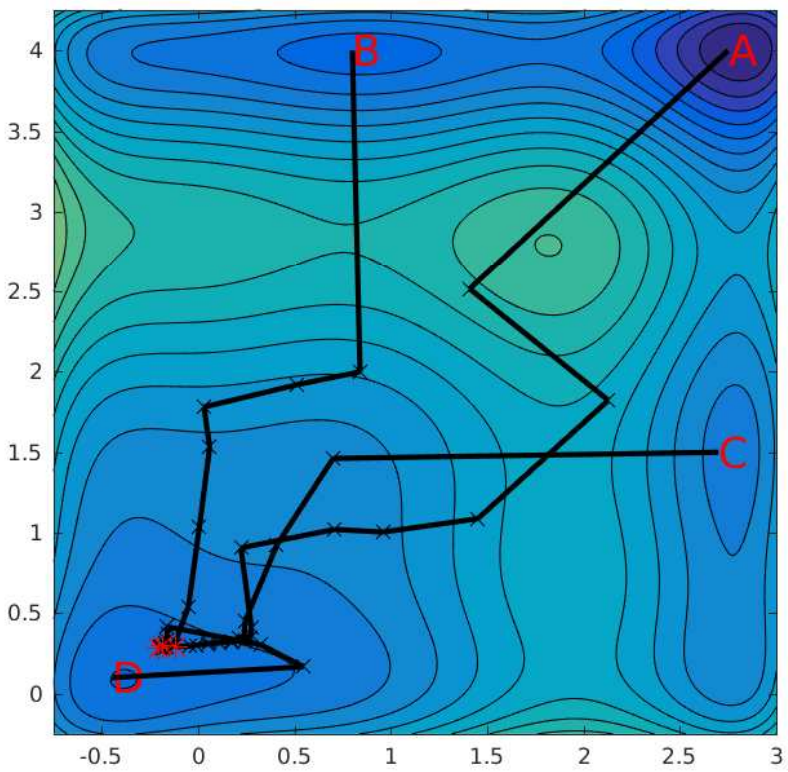


$\Psi_{\mathcal{U}_{0.5}}(x)$

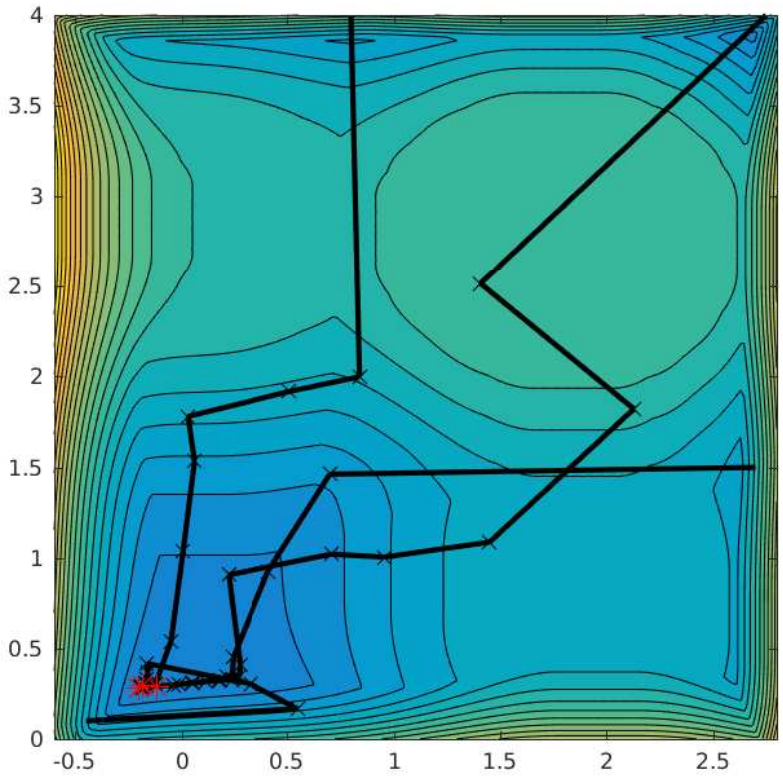
- Recover global within 250 total evaluations
- $\mathfrak{U}^0 = \{\pm 0.5 e_i : i = 1, 2\}$

See also [Bertsimas, Nohadani; JOGO 2010]; [Conn, Vicente; OMS 2012]

DFOA Trajectories Bertsimas-Nohadani-Teo



$g(x) = f(x, 0)$



$\Psi_{\mathcal{U}_{0.5}}(x)$

- Recover global within 250 total evaluations
- $\mathfrak{U}^0 = \{\pm 0.5 e_i : i = 1, 2\}$

See also [Bertsimas, Nohadani; JOGO 2010]; [Conn, Vicente; OMS 2012]

Goldfarb Biquadratics

$$\min_{x\in\mathbb{R}^n} \Psi_{\mathcal{U}_\alpha}(x) := \min_{x\in\mathbb{R}^n} \max_{(L,b)\in\mathcal{U}_\alpha} \frac{1}{2}x^\top L^\top Lx + b^\top x,$$

Uncertainty set $\mathcal{U}_\alpha$ $(\alpha \geq 0)$:

$$\mathcal{U}_\alpha := \left\{(L,b) \in \mathbb{L}^n \times \mathbb{R}^n : \ |L_{ij} - \hat{L}_{ij}| \leq \alpha, \forall i \geq j; \ |b_i - \hat{b}_i| \leq \alpha, \forall i\right\}$$

Nominal $\hat{L} \in \mathbb{L}^n$ lower triangular with nonzero diagonal entries; $\hat{b} \in \mathbb{R}^n$

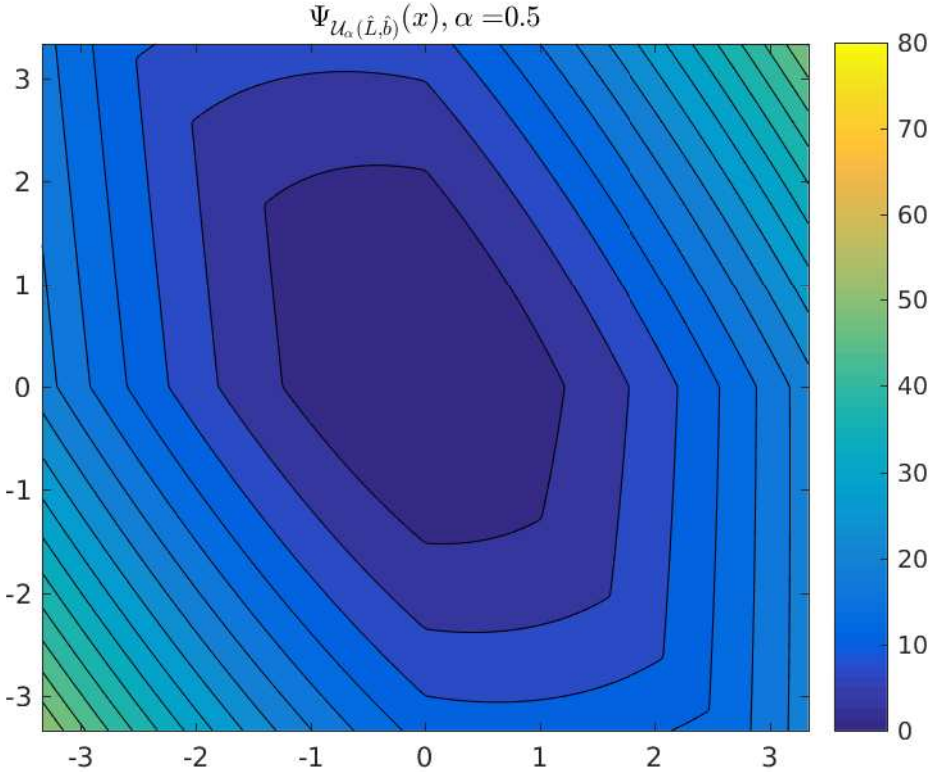
Goldfarb Biquadratics

$$\min_{x\in\mathbb{R}^n} \Psi_{\mathcal{U}_\alpha}(x) := \min_{x\in\mathbb{R}^n} \max_{(L,b)\in\mathcal{U}_\alpha} \frac{1}{2}x^\top L^\top Lx + b^\top x,$$

Uncertainty set $\mathcal{U}_\alpha$ $(\alpha \geq 0)$:

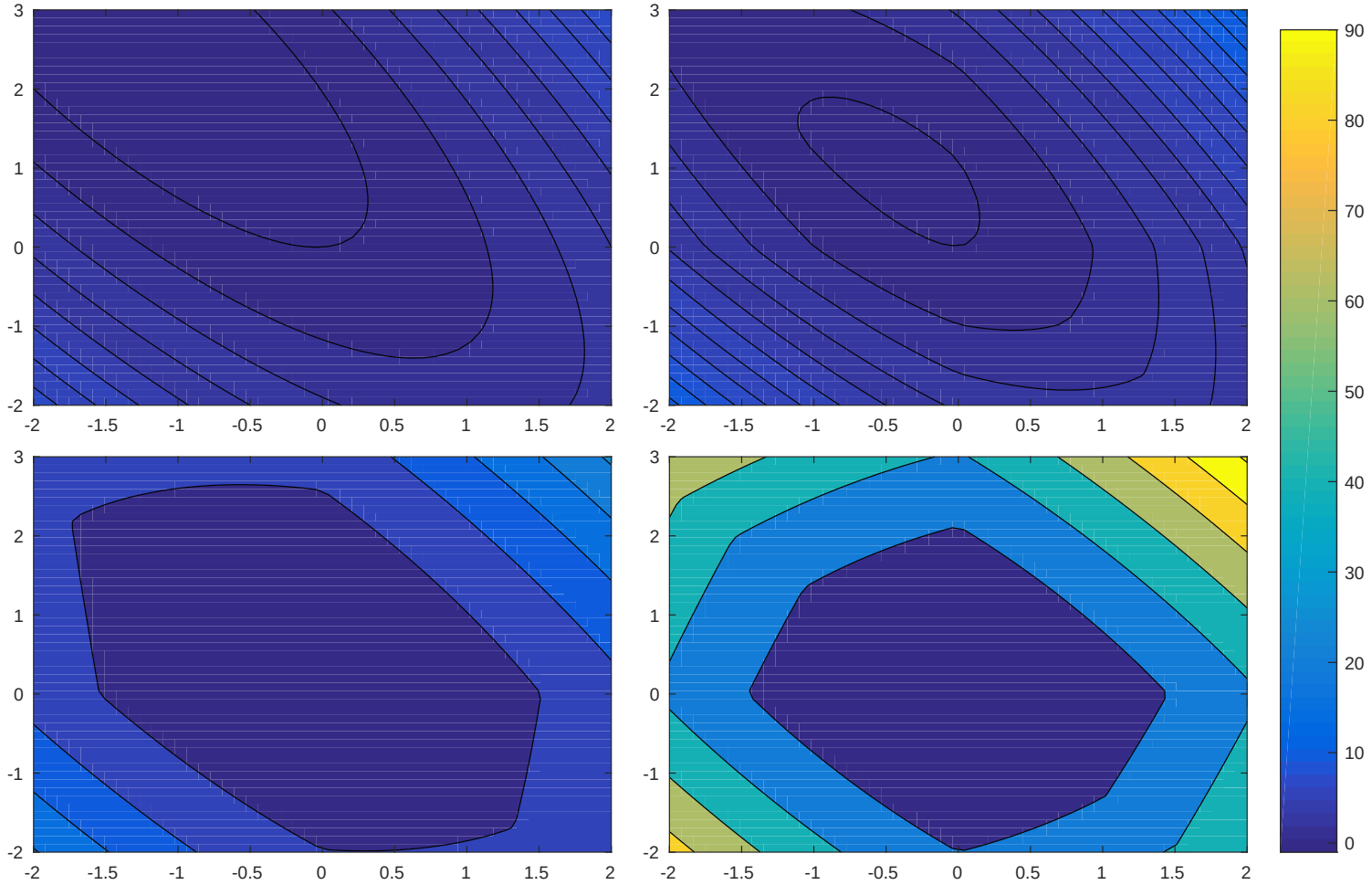
$$\mathcal{U}_\alpha := \left\{(L,b) \in \mathbb{L}^n \times \mathbb{R}^n : \ |L_{ij} - \hat{L}_{ij}| \leq \alpha, \forall i \geq j; \ |b_i - \hat{b}_i| \leq \alpha, \forall i\right\}$$

Nominal $\hat{L} \in \mathbb{L}^n$ lower triangular with nonzero diagonal entries; $\hat{b} \in \mathbb{R}^n$



Goldfarb Biquadratics Example (Varying α)

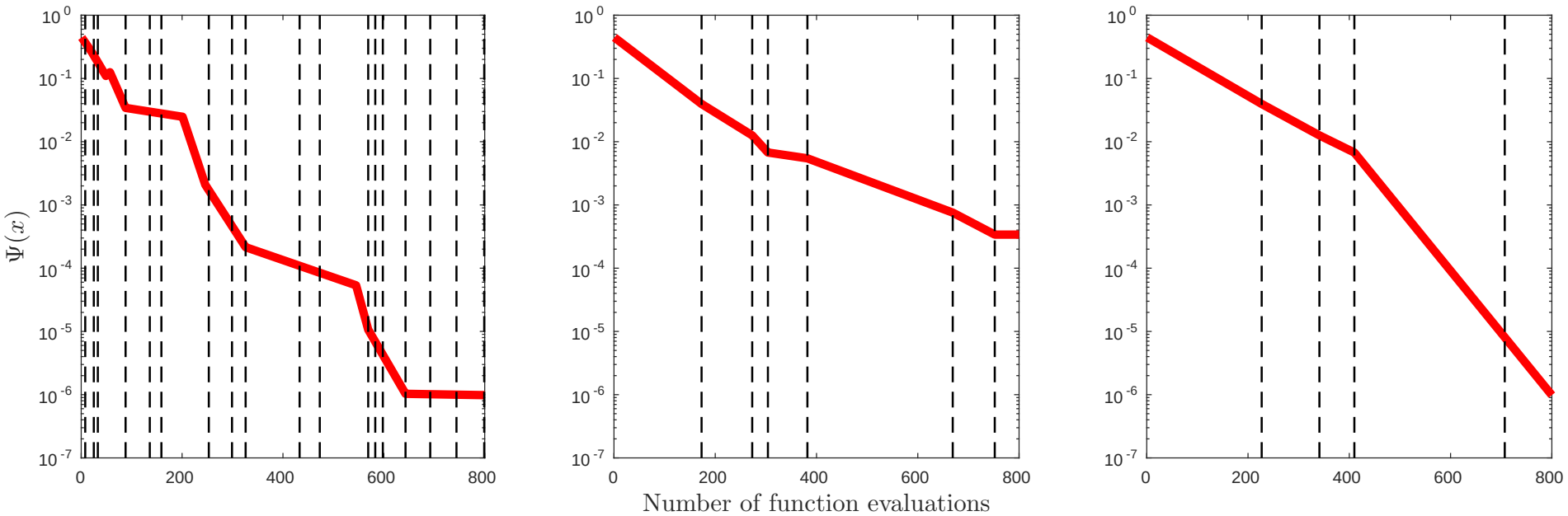
$\Psi_{\mathcal{U}_\alpha}(x)$ for a randomly generated set of nominal $\hat{L}, \hat{u}$



$$\alpha = \begin{array}{c|c} 0 & 0.125 \\ \hline 0.5 & 2 \end{array}$$

RCQP Typical Results

$\Psi(x)$ progress; dashed lines indicate the end of a phase 1



Gaussian RBF | Uniform random sampling | Optimal phase 2

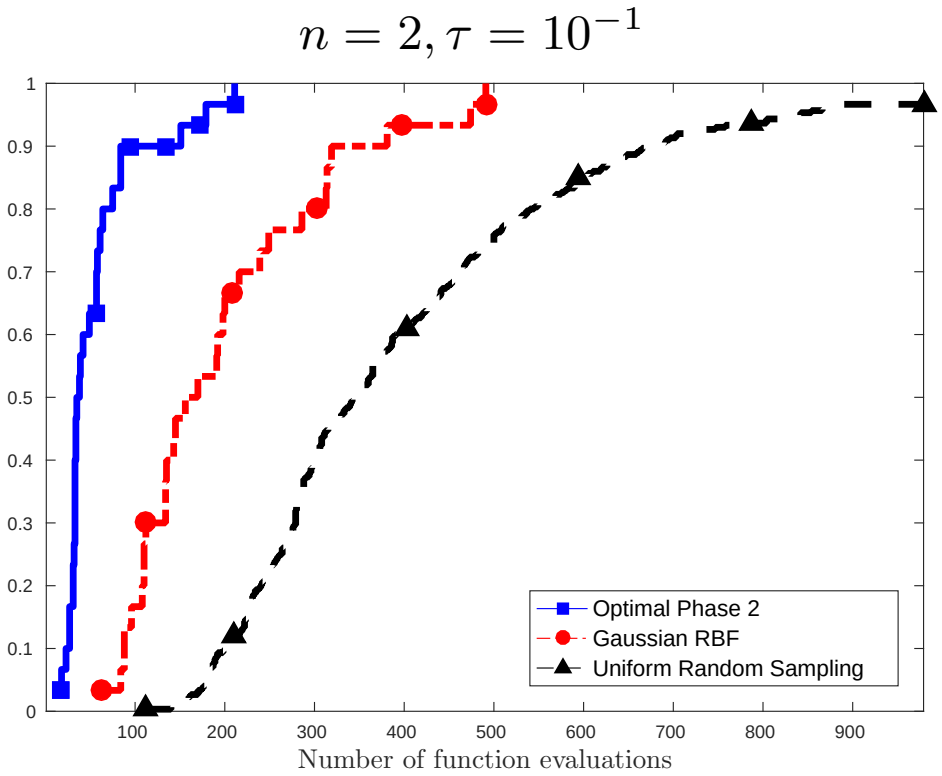
RBF $m^f(u)$ interpolating on available points $\{(x^{k+1}, u)\}$
Obtain u' from approximate solution of $\max_{u \in \mathcal{U}} m^f(u)$

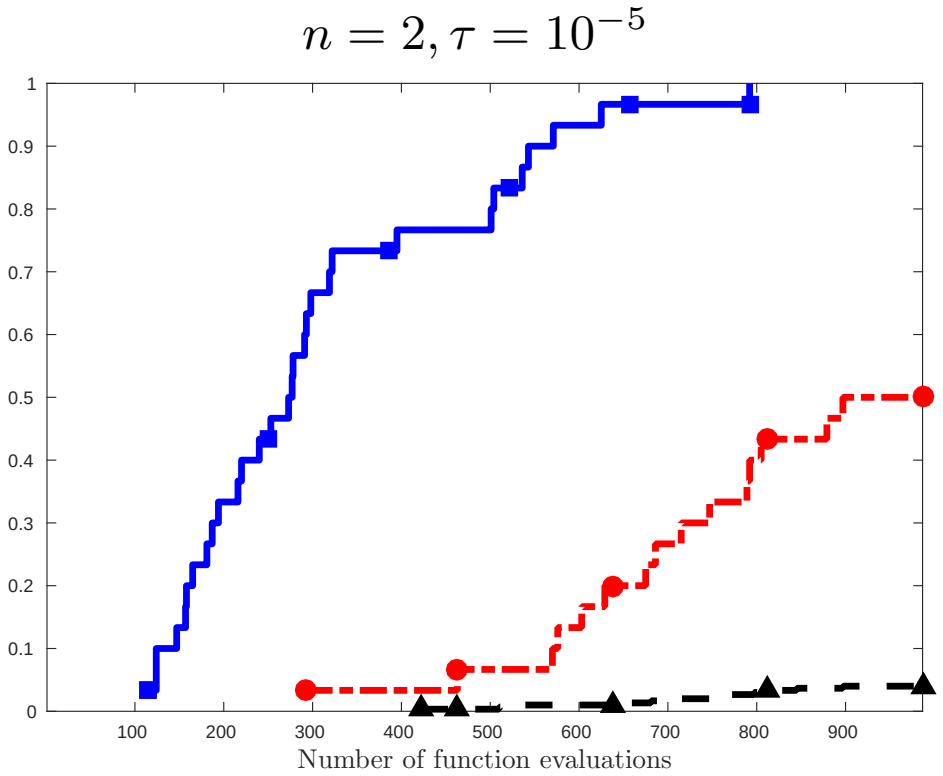
Phase 2:

Uniform Ω^k from $\lceil \beta m \rceil$ Unif($\mathcal{U}$) samples
$u' \in \arg\max_{u \in \Omega^k} f(x^{k+1}, u)$

Optimal $u' \in \arg\max_{u \in \mathcal{U}} f(x^{k+1}, u)$

Ψ Data Profiles for RCQP

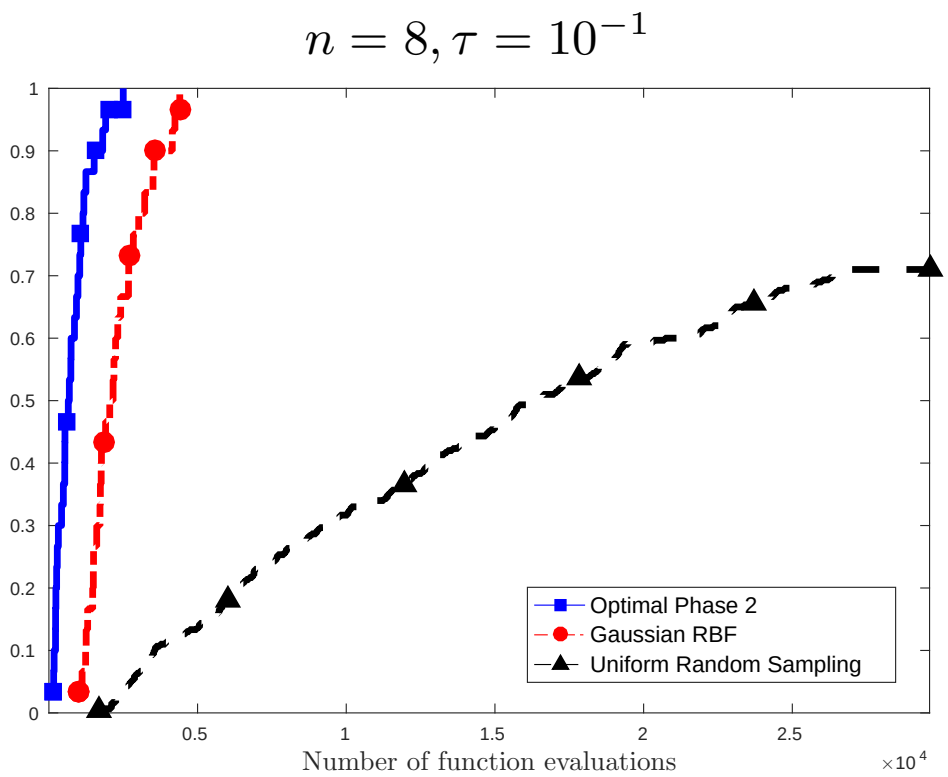


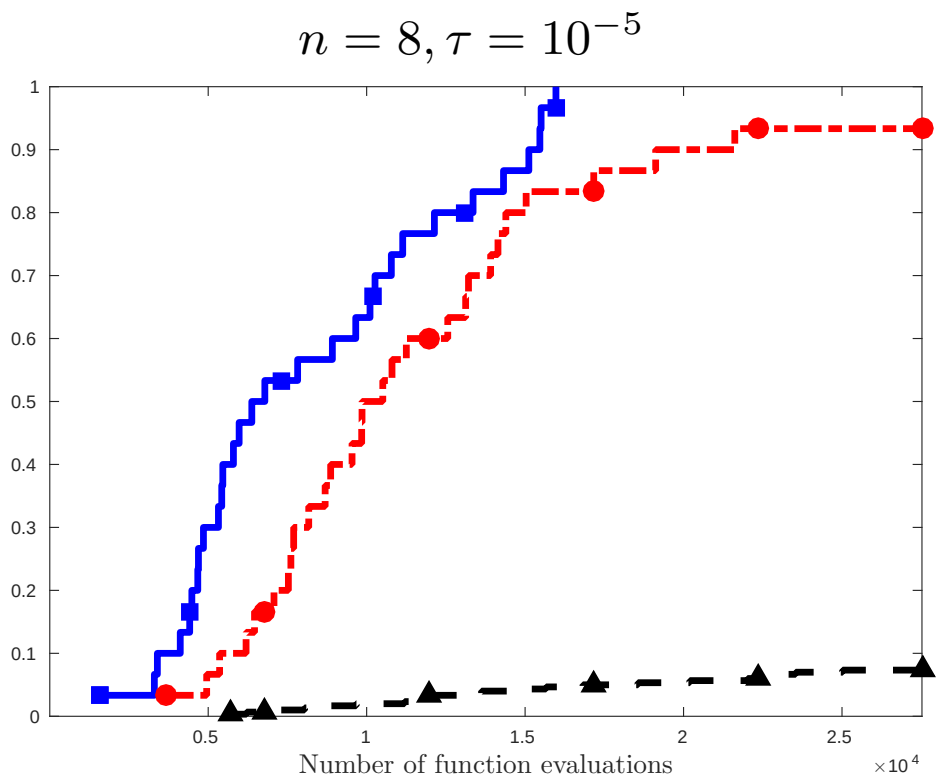


30 random RCQPs (many random trials for uniform sampling)

$$\Psi_{\mathcal{U}}(x^0) - \Psi_{\mathcal{U}}(x) \geq (1-\tau)\left(\Psi_{\mathcal{U}}(x^0) - \Psi_{\mathcal{U}}(x^{\text{best}})\right)$$

Ψ Data Profiles for RCQP





30 random RCQPs (many random trials for uniform sampling)

$$\Psi_{\mathcal{U}}(x^0) - \Psi_{\mathcal{U}}(x) \geq (1-\tau)\left(\Psi_{\mathcal{U}}(x^0) - \Psi_{\mathcal{U}}(x^{\text{best}})\right)$$

Grey-box algorithms for optimization problems with black-box components

Model-based outer approximation looks promising [Menickelly, W.; Prep. 2017]

- Employs framework of Polak's inexact outer approximation
- Builds smooth models with an inner trust-region approach
- Uses manifold sampling for composite nonsmooth $h(S(x)) = \Psi_{\mathcal{U}}(x; S(x))$
  - $h = \|\cdot\|_1$ [Larson, Menickelly, W.; SIOPT 2016]; h pl [Khan, Larson, W.; Prep. 2017]
- Interested in exploiting implementation error structure $f(x, u) = g(x + u)$
  - If $g(y)$ available, then $f(y, u - y) = f(y - u, u) = g(y)$ available for all u

← **Matt Menickelly**
& the NRO gang:
Sven Leyffer
Todd Munson
Charlie Vanaret

www.mcs.anl.gov/~wild

Thank YOU!