NONSMOOTH OPTIMIZATION: THINKING OUTSIDE OF THE BLACK BOX

Claudia Sagastizábal

Eletrobras Cepel

mailto:sagastiz@impa.br, http://www.impa.br/~sagastiz

SIAM Conference on Optimization,
Darmstadt, May 18th 2011

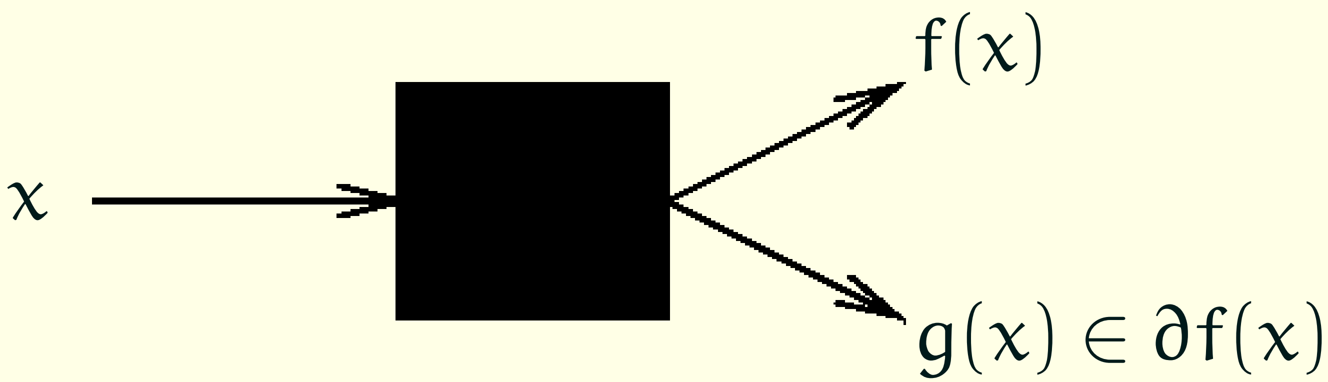
With thanks to:
AFOSR Grant FA9550-08-1-0370, NSF Grant DMS 0707205,
and CNPq & Faperj from Brazil

NSO Algorithms

For a convex nonsmooth function, solving

$$\boxed{\min f(x)}$$

with a black box method



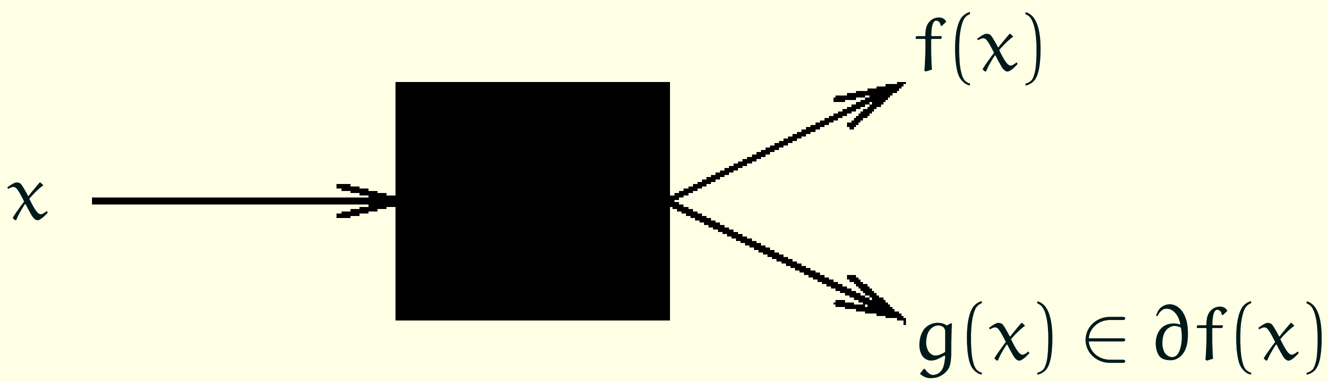
is doomed to slow convergence speed.

NSO Algorithms

For a convex nonsmooth function, solving

$$\min f(x)$$

with a black box method



is doomed to slow convergence speed.

Better performance possible by exploiting structure

How does structure appear?

– Explicitly

as a sum

as a composition

– Implicitly

U-Lagrangian

VU-decomposition

partly smooth functions

How does structure appear?

– Explicitly

as a sum

as a composition

– Implicitly

U-Lagrangian

VU-decomposition

partly smooth functions

How does structure appear?

– Explicitly

as a sum

as a composition

$\neq$ **black boxes**

– Implicitly

U-Lagrangian

VU-decomposition

partly smooth functions

How does structure appear?

– Explicitly

as a sum

as a composition

$\neq$ **black boxes**

– Implicitly

U-Lagrangian

VU-decomposition

partly smooth functions

digging tools

Explicit Structure:

Opening the Black Box

A convex partly nonsmooth function

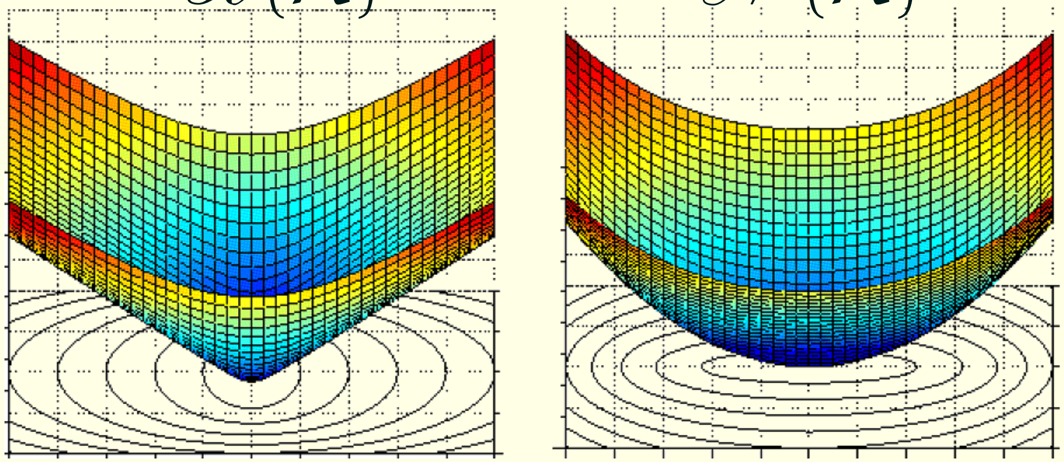
For $x \in \mathbb{R}^n$, given matrices $A \succeq 0$, $B \succ 0$,

$$f(x) = \sqrt{x^\top A x} + x^\top B x$$

has a unique minimizer at 0.

On $\mathscr{N}(A)$ the function is not differentiable, and the first term vanishes: $f|_{\mathscr{N}(A)}$ looks smooth.

$\mathscr{R}(A)$ $\qquad$ $\mathscr{N}(A)$



This function has several interesting structures

If no structure at all

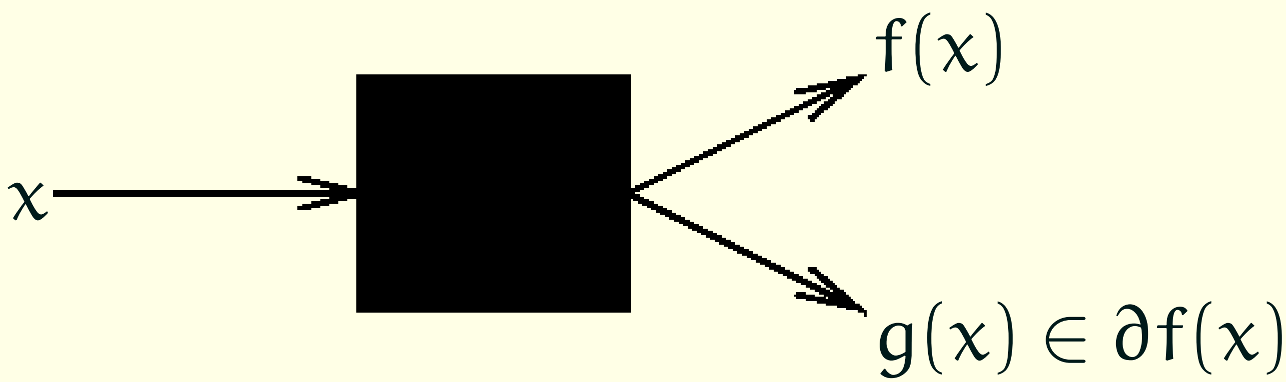
$$f(x) = \sqrt{x^\top A x} + x^\top B x$$

This function has several interesting structures

If no structure at all

$$f(x) = \sqrt{x^\top A x} + x^\top B x$$

This defines the **black box**:



This function has several interesting structures

Sum structure

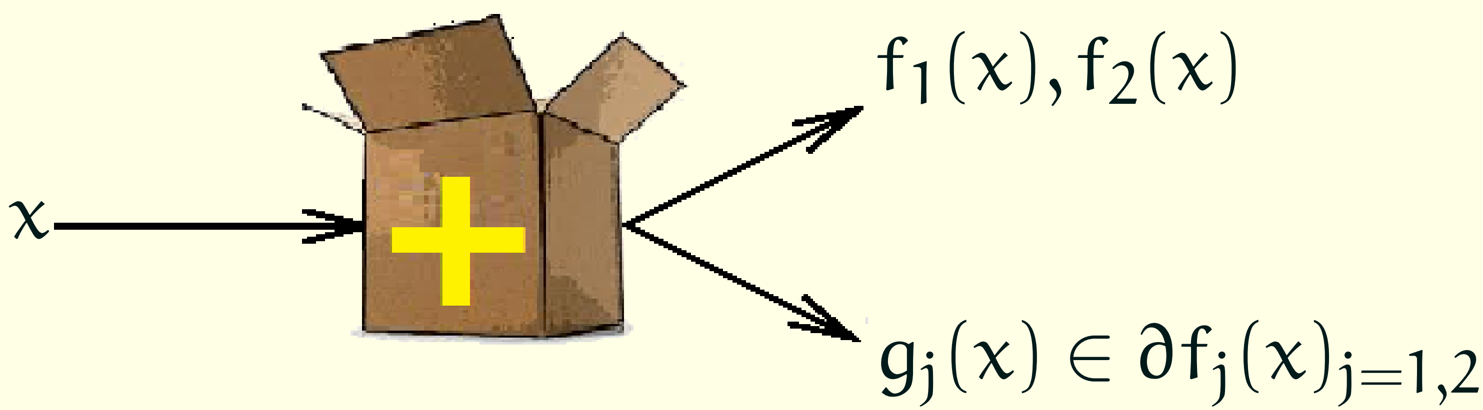
$$f(x) = f_1(x) + f_2(x) \text{ with } \begin{cases} f_1(x) = \sqrt{x^\top A x} \\ f_2(x) = x^\top B x \end{cases}$$

This function has several interesting structures

Sum structure

$$f(x) = f_1(x) + f_2(x) \text{ with } \begin{cases} f_1(x) = \sqrt{x^\top A x} \\ f_2(x) = x^\top B x \end{cases}$$

This defines a **sum black box**:



This function has several interesting structures

Composite structure

$$f(x) = (h \circ c)(x) \text{ with } \begin{cases} c(x) = \left(x, x^\top B x\right) \in \mathbb{R}^{n+1} \\ h(C) = \sqrt{C_{1:n}^\top A C_{1:n}} + C_{n+1} \end{cases}$$

for C smooth and h positively homogeneous

This function has several interesting structures

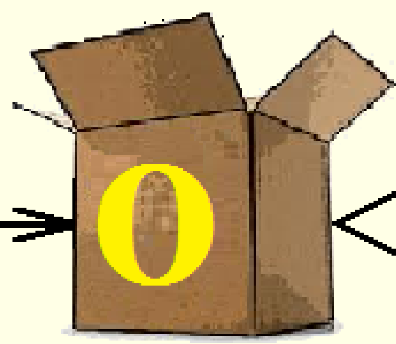
Composite structure

$$f(x) = (h \circ c)(x) \text{ with } \begin{cases} c(x) = \left(x, x^\top B x\right) \in \mathbb{R}^{n+1} \\ h(C) = \sqrt{C_{1:n}^\top A C_{1:n}} + C_{n+1} \end{cases}$$

for C smooth and h positively homogeneous

This defines a **composite black box**:

x



$C := c(x)$ and $h(C)$

Jacobian $Dc(x)$ and

$G(C) \in \partial h(C)$

This function has several interesting structures

Missing information structure

Suppose not all of A/B is known/accessible,

so that only **estimates** are available for f

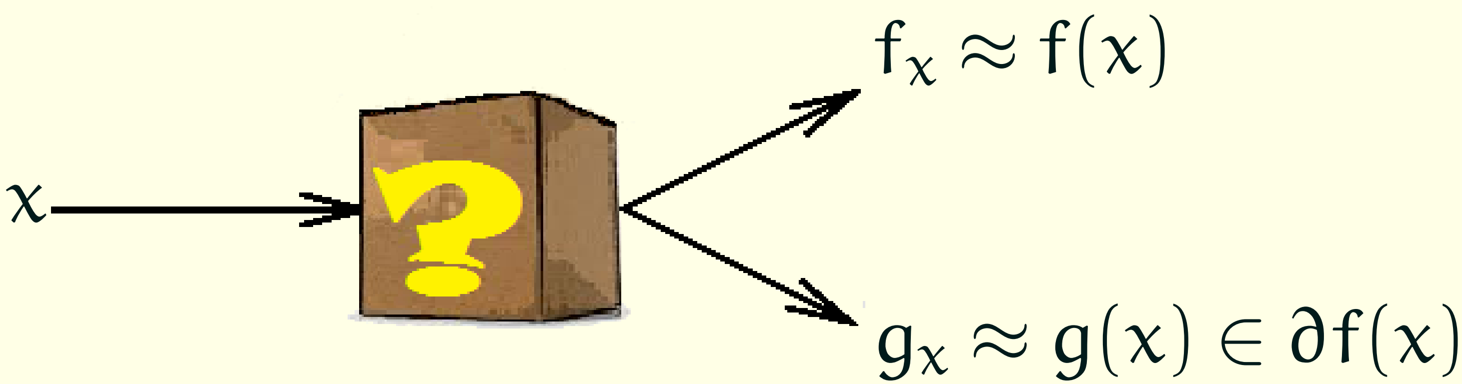
This function has several interesting structures

Missing information structure

Suppose not all of A/B is known/accessible,

so that only **estimates** are available for f

This defines a **noisy black box**:



How to use explicit structure in an algorithm?

Black box information defines **pieces** that put together create a **model φ** of the function f.

The model is used to define iterates not too far away from a "good" past iterate, $\hat{x}$. At iteration i,

$$x^{i+1} \text{ minimizes } \varphi(x) + \frac{1}{2}\mu|x - \hat{x}|^2$$

How to use explicit structure in an algorithm?

Black box information defines **pieces** that put together create a **model φ** of the function f.

The model is used to define iterates not too far away from a "good" past iterate, $\hat{x}$. At iteration i,

$$x^{i+1} \text{ minimizes } \varphi(x) + \frac{1}{2}\mu|x - \hat{x}|^2$$

"pieces" chosen to make minimization simple (QP)

for example, "piece"=linearization:

$$x^i \longrightarrow \blacksquare \begin{cases} f^i = f(x^i) \\ g^i = g(x^i) \end{cases} \Longrightarrow \qquad f^i + g^{i\top}(x - x^i)$$

How to use explicit structure in an algorithm?

Black box information defines **pieces** that put together create a **model φ** of the function f.

The model is used to define iterates not too far away from a "good" past iterate, $\hat{x}$. At iteration i,

$$x^{i+1} \text{ minimizes } \varphi(x) + \frac{1}{2}\mu|x - \hat{x}|^2$$

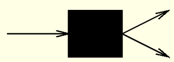
"pieces" chosen to make minimization simple (QP)

for example, "piece"=linearization:

$$x^i \longrightarrow \blacksquare \begin{array}{l} \nearrow f^i = f(x^i) \\ \searrow g^i = g(x^i) \end{array} \qquad \Longrightarrow \varphi(x) = \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\}$$

$$x^{i+1} = \arg\min_x \varphi(x) + \frac{1}{2}\mu|x - \hat{x}|^2$$

for example, "piece"=linearization:

x^i  $f^i = f(x^i)$

$g^i = g(x^i)$

$$\Longrightarrow \varphi(x) = \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\}$$

Some jargon

$\hat{x}$ is a serious point

$\bigcup_i (x^i, f^i, g^i)$ is the bundle $\mathscr{B}$

If x^{i+1} gives sufficient decrease for f, it becomes the next $\hat{x}$

Otherwise, it is declared a null point

$$x^{i+1} = \arg\min_x \varphi(x) + \frac{1}{2}\mu|x - \hat{x}|^2$$

for example, “piece”=linearization:

$$\boxed{x^i \longrightarrow \blacksquare \begin{array}{l} f^i = f(x^i) \\ g^i = g(x^i) \end{array}} \implies \varphi(x) = \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\}$$

Some jargon

$\hat{x}$ is a <u>serious point</u>

$\bigcup_i (x^i, f^i, g^i)$ is the <u>bundle</u> $\mathscr{B}$

If x^{i+1} gives sufficient decrease for f, it becomes the next $\hat{x}$

Otherwise, it is declared a <u>null</u> point

$$x^{i+1} = \arg\min_x \varphi(x) + \frac{1}{2}\mu|x-\hat{x}|^2$$

for example, “piece”=linearization:

$$x^i \longrightarrow \blacksquare \begin{array}{l} f^i = f(x^i) \\ g^i = g(x^i) \end{array} \quad \Longrightarrow \varphi(x) = \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\}$$

different boxes

$\uparrow$ **different models**

Some jargon

$\hat{x}$ is a serious point

$\bigcup_i (x^i, f^i, g^i)$ is the bundle $\mathscr{B}$

If x^{i+1} gives sufficient decrease for f, it becomes the next $\hat{x}$

Otherwise, it is declared a null point

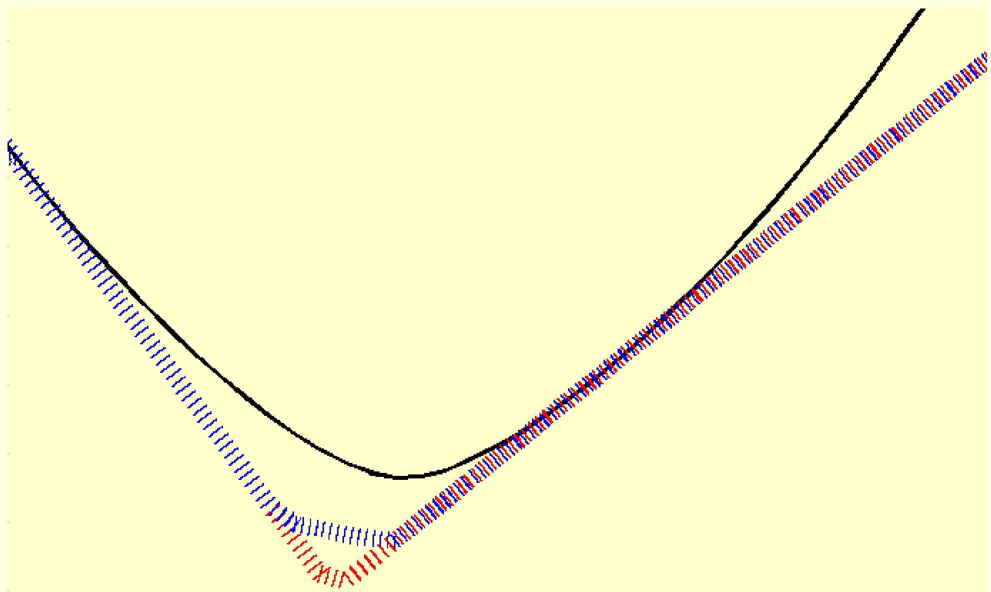
Structured models for f

No structure

$$\begin{aligned}\varphi(x) &= \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\} \\ &= \max_i \left\{ (f_1^i + f_2^i) + (g_1^i + g_2^i)^\top (x - x^i) \right\}\end{aligned}$$

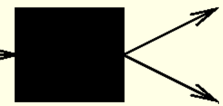
Sum structure

$$\begin{aligned}\varphi(x) &= \max_i \left\{ f_1^i + g_1^{i\top}(x - x^i) \right\} \\ &+ \max_i \left\{ f_2^i + g_2^{i\top}(x - x^i) \right\}\end{aligned}$$



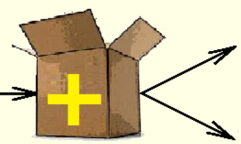
Structured models for f

No structure



$$\begin{aligned}\varphi(x) &= \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\} \\ &= \max_i \left\{ (f_1^i + f_2^i) + (g_1^i + g_2^i)^\top (x - x^i) \right\}\end{aligned}$$

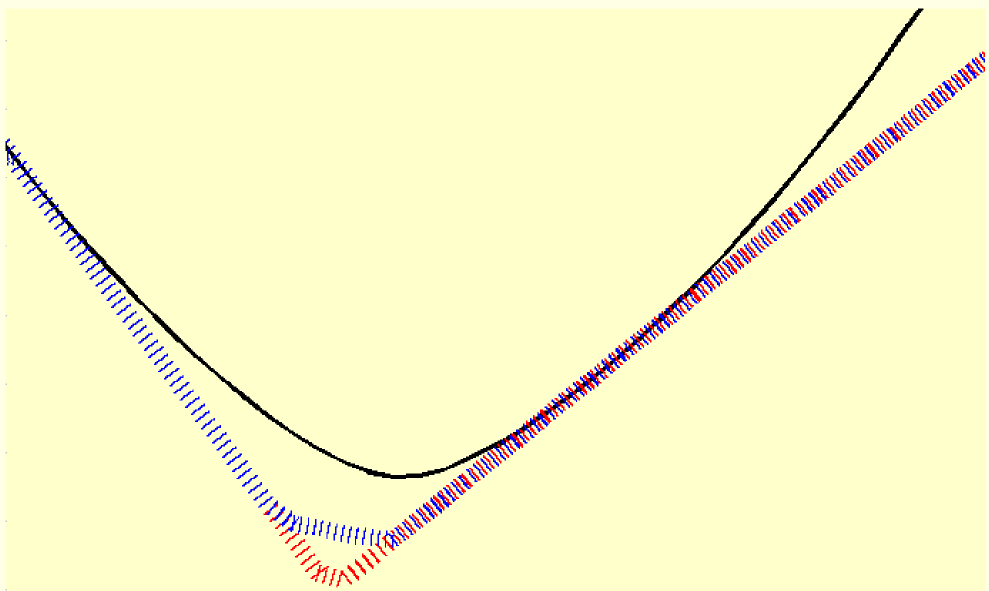
Sum structure



$$\begin{aligned}\varphi(x) &= \max_i \left\{ f_1^i + g_1^{i\top}(x - x^i) \right\} \\ &+ \max_i \left\{ f_2^i + g_2^{i\top}(x - x^i) \right\}\end{aligned}$$

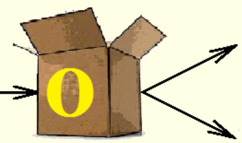
Larger

QP



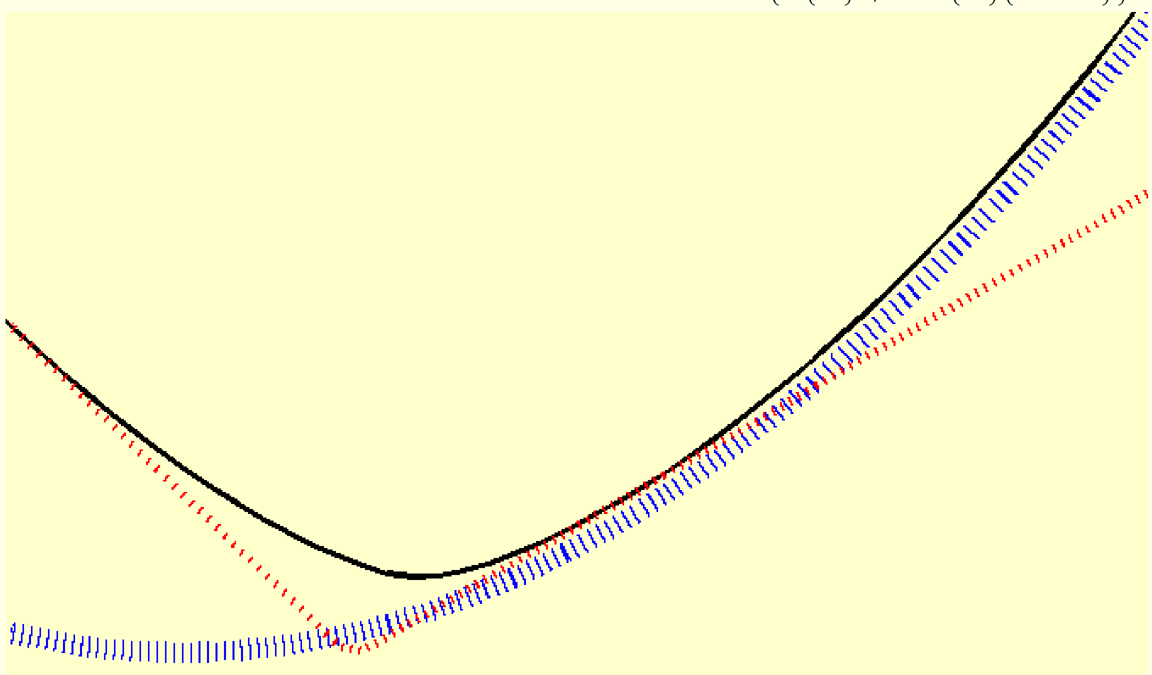
Structured models for f

Composite structure



$$\varphi(x) = \max_i \left\{ G^{i\top} \Big(c(\hat{x}) + Dc(\hat{x})(x - \hat{x}) \Big) \right\}$$

$$\approx h(c(\hat{x}) + Dc(\hat{x})(x - \hat{x}))$$

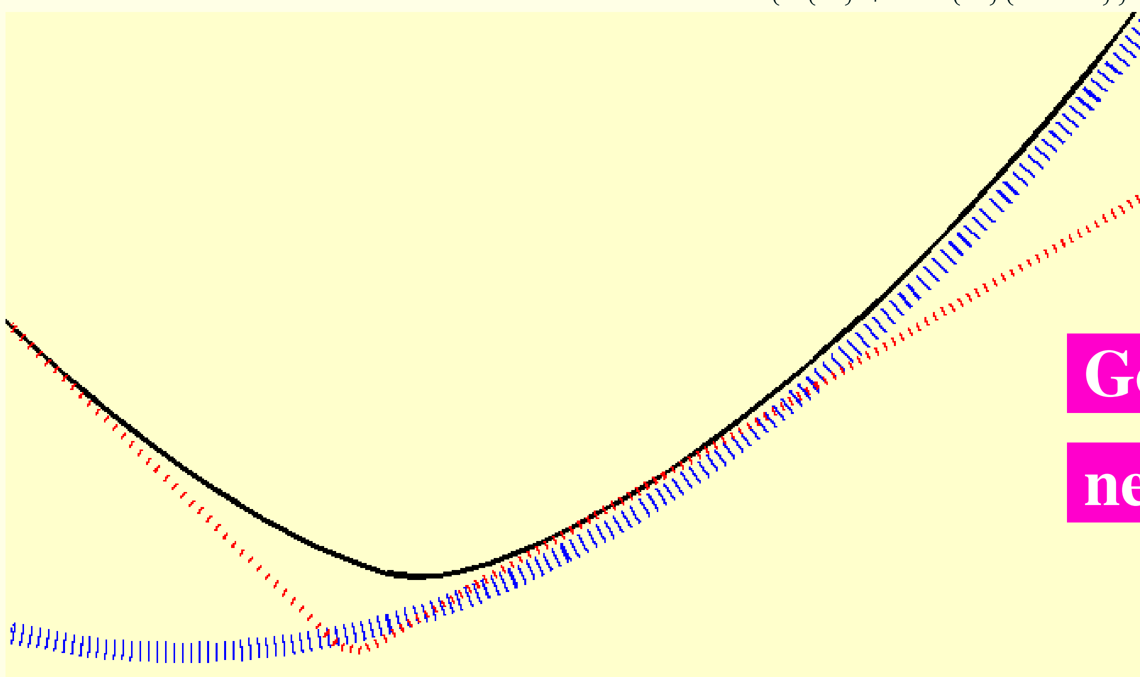


Structured models for f

Composite structure

$$\varphi(x) = \max_i \left\{ G^{i\top} \Big(c(\hat{x}) + Dc(\hat{x})(x - \hat{x}) \Big) \right\}$$

$$\approx h(c(\hat{x}) + Dc(\hat{x})(x - \hat{x}))$$

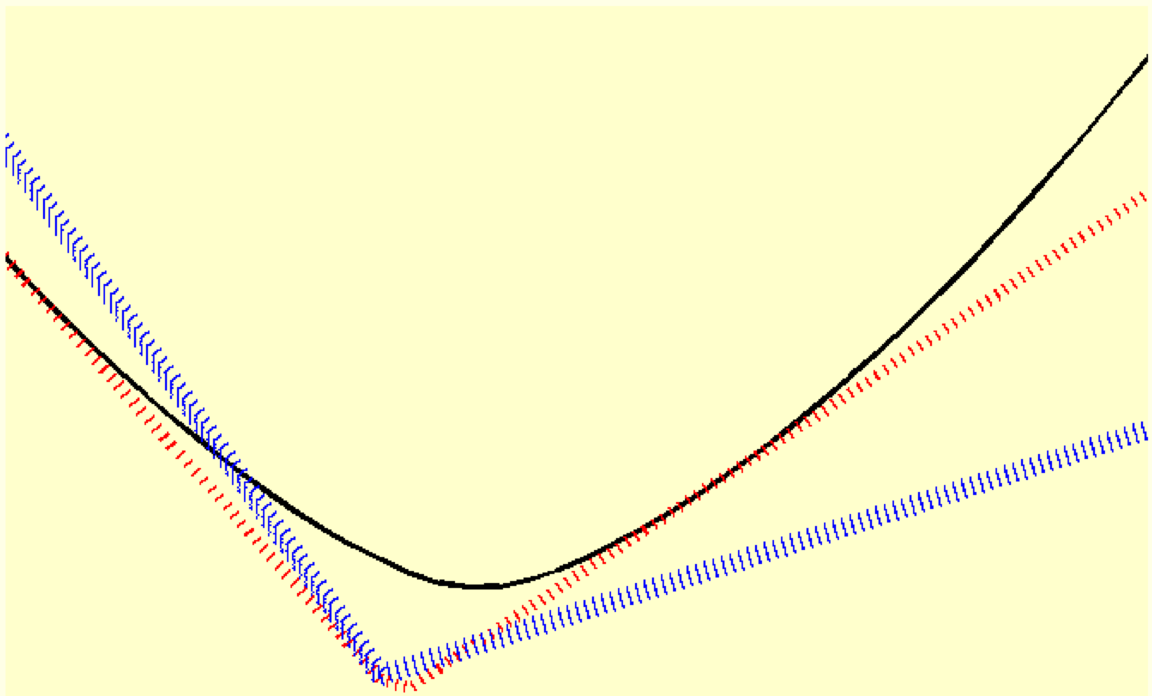


Good

near $\hat{x}$

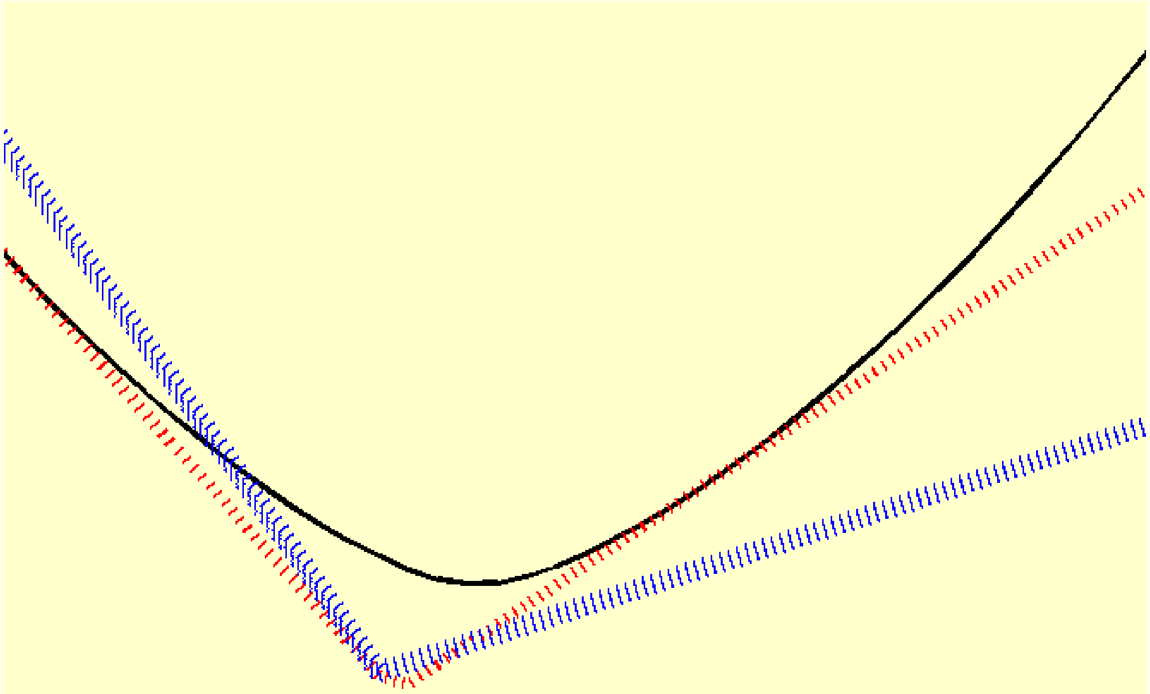
Inexact models for f

Missing structure  $\varphi(x) = \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\}$



Inexact models for f

Missing structure  $\varphi(x) = \max_i \left\{ f^i + g^{i\top}(x - x^i) \right\}$

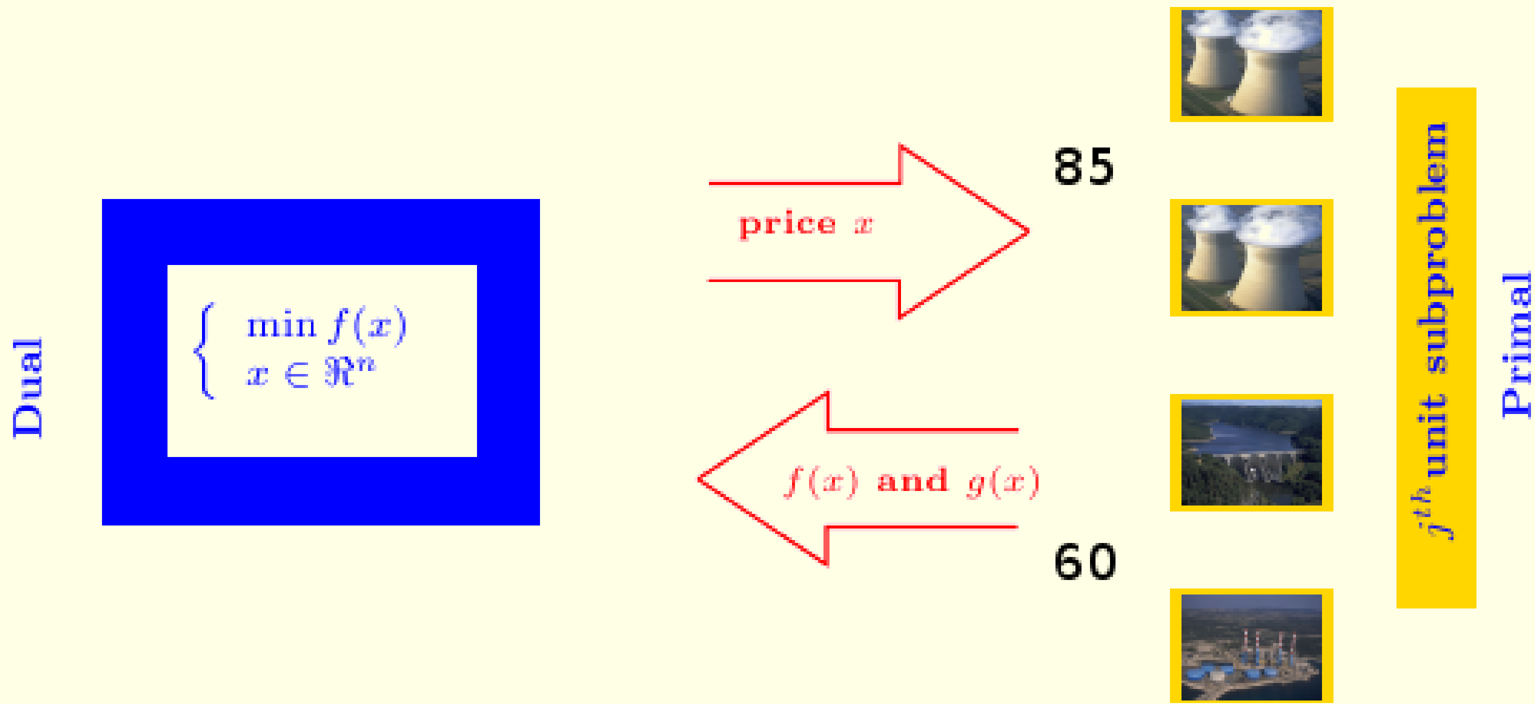


φ may cut f

excessive noise is attenuated via μ

Stochastic Programming Applications

Mid-term planning for power generation

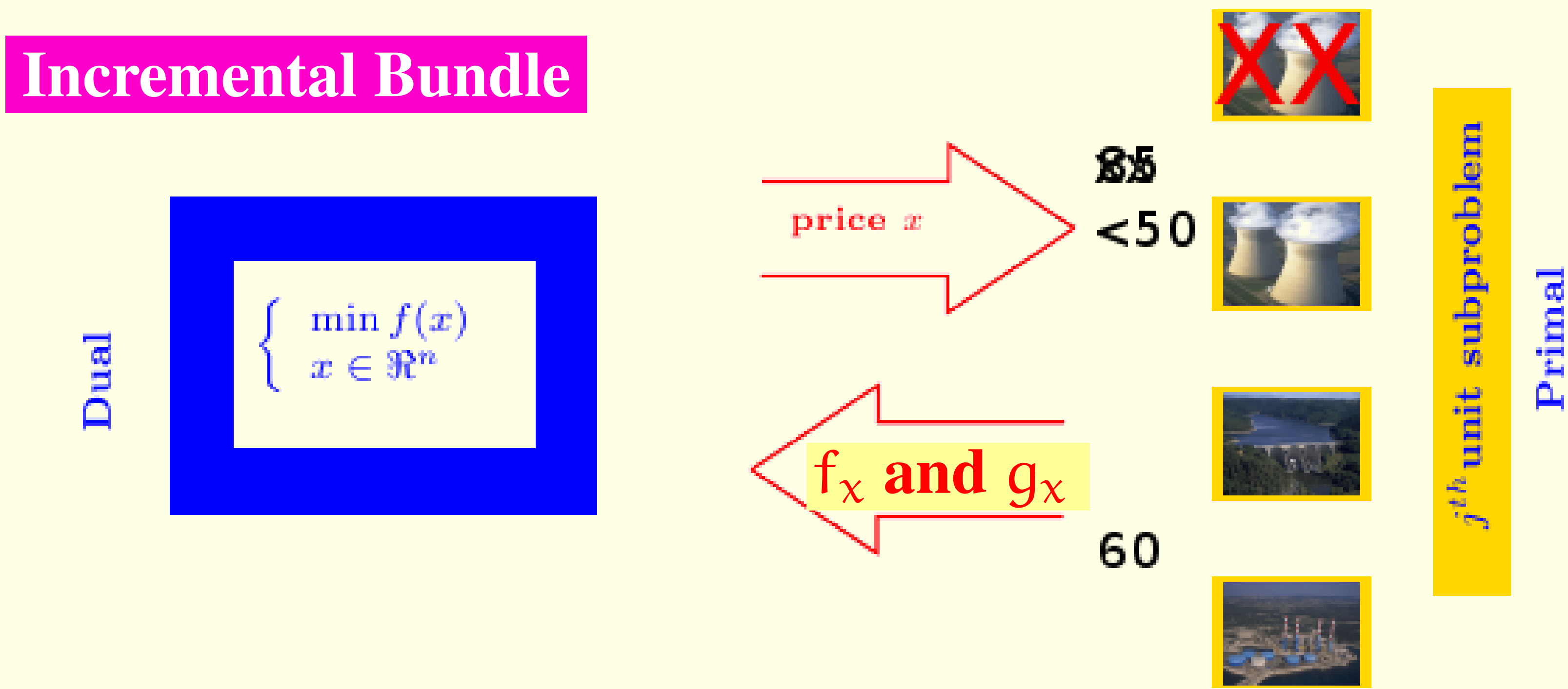


Scenario tree with 50,000 nodes

Nuclear LPs with 100,000 variables and 300,000 constraints

Stochastic Programming Applications

Mid-term planning for power generation

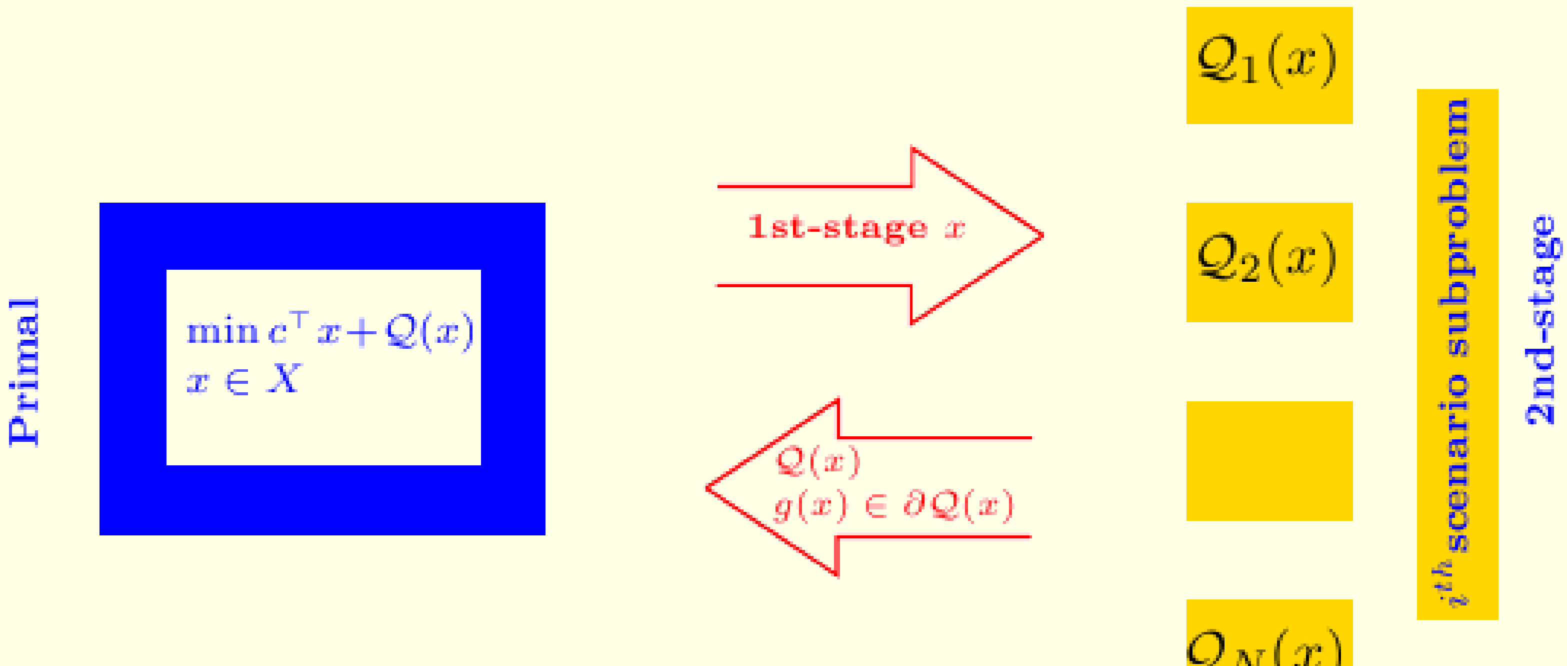


Skips Nuclear LPs (alternating) $\equiv$ noisy black box

25% less CPU time than exact bundle, same accuracy

Stochastic Programming Applications

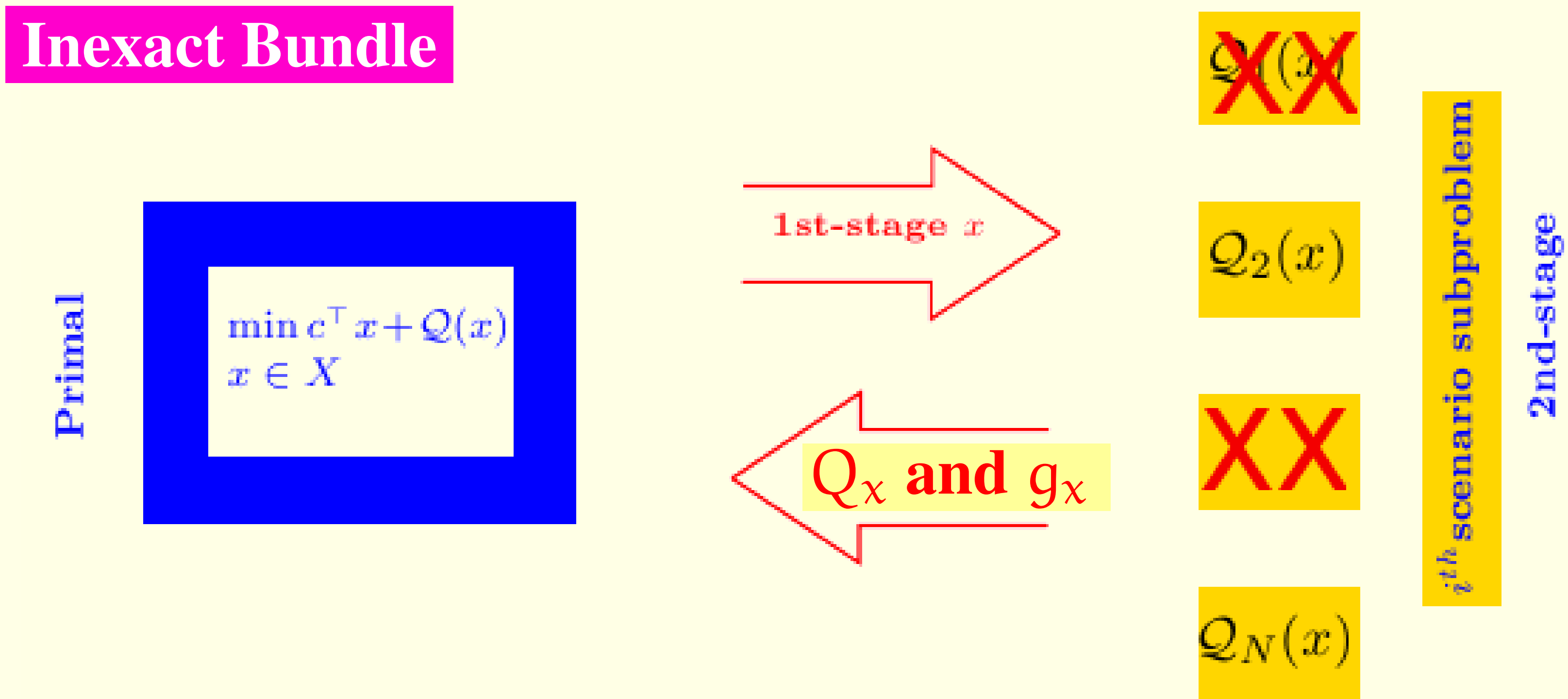
2-stage stochastic linear programs



L-shaped decomposition into N scenarios

Stochastic Programming Applications

2-stage stochastic linear programs

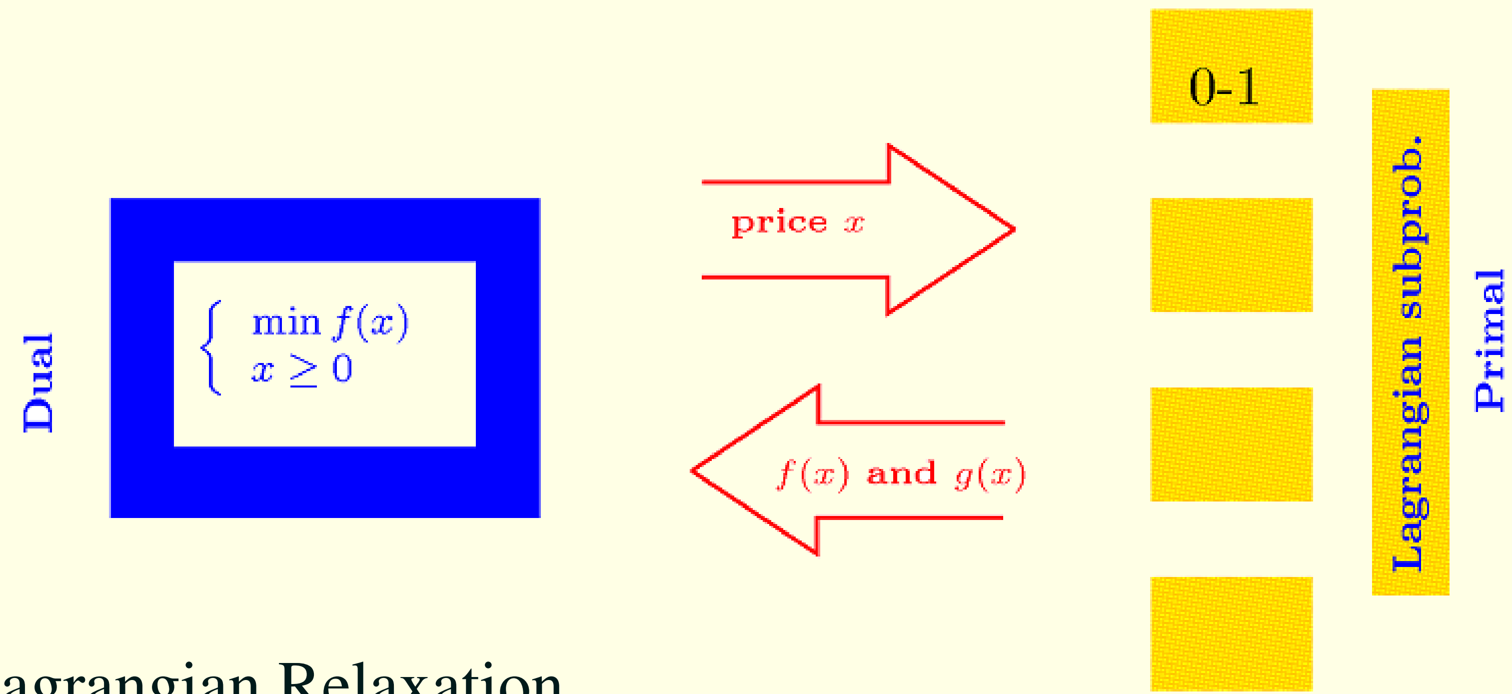


Skips 80% LPs solution $\equiv$ noisy black box

4 times faster than L-shaped, same accuracy

Combinatorial Optimization Applications

Exponential number of hard constraints

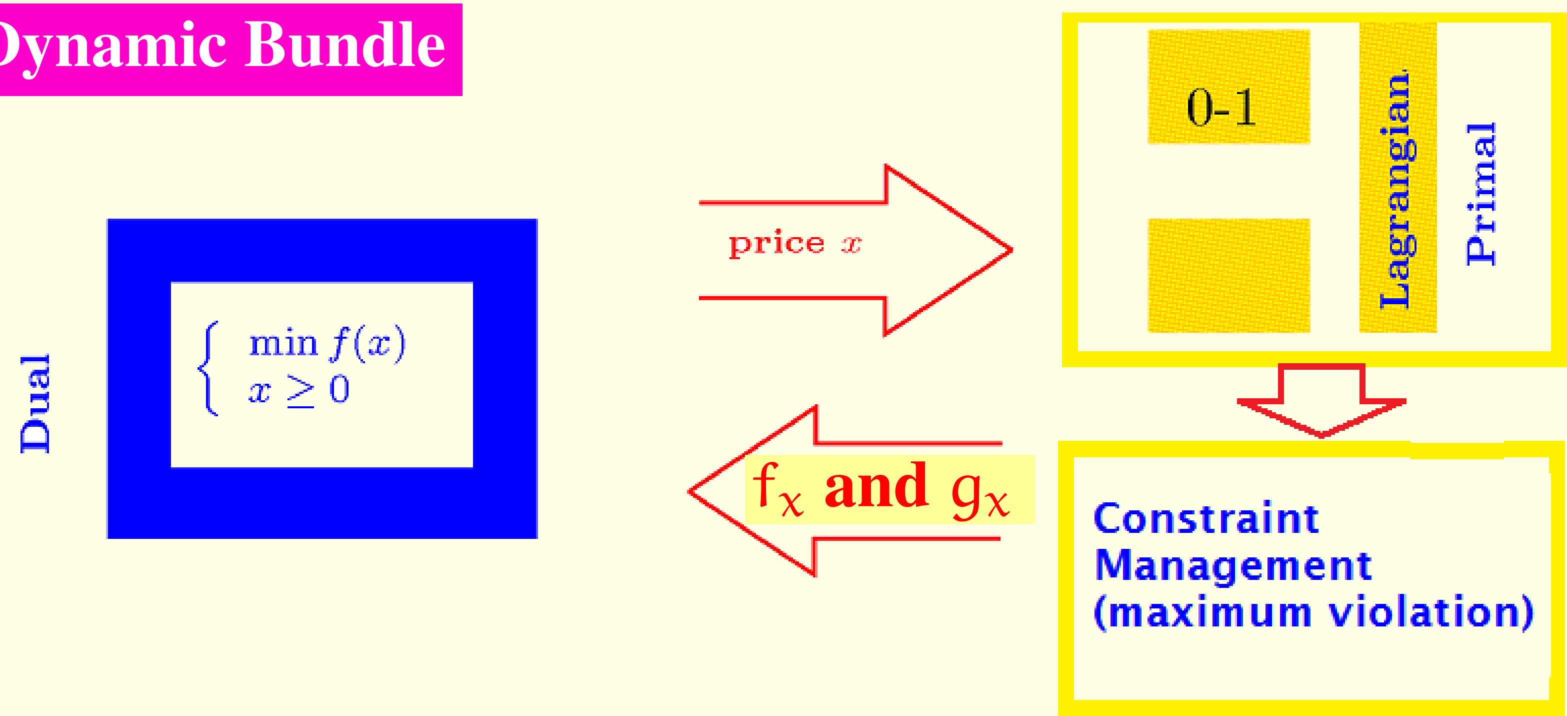


Lagrangian Relaxation

Combinatorial Optimization Applications

Exponential number of hard constraints

Dynamic Bundle



Like "Relax-and-cut"

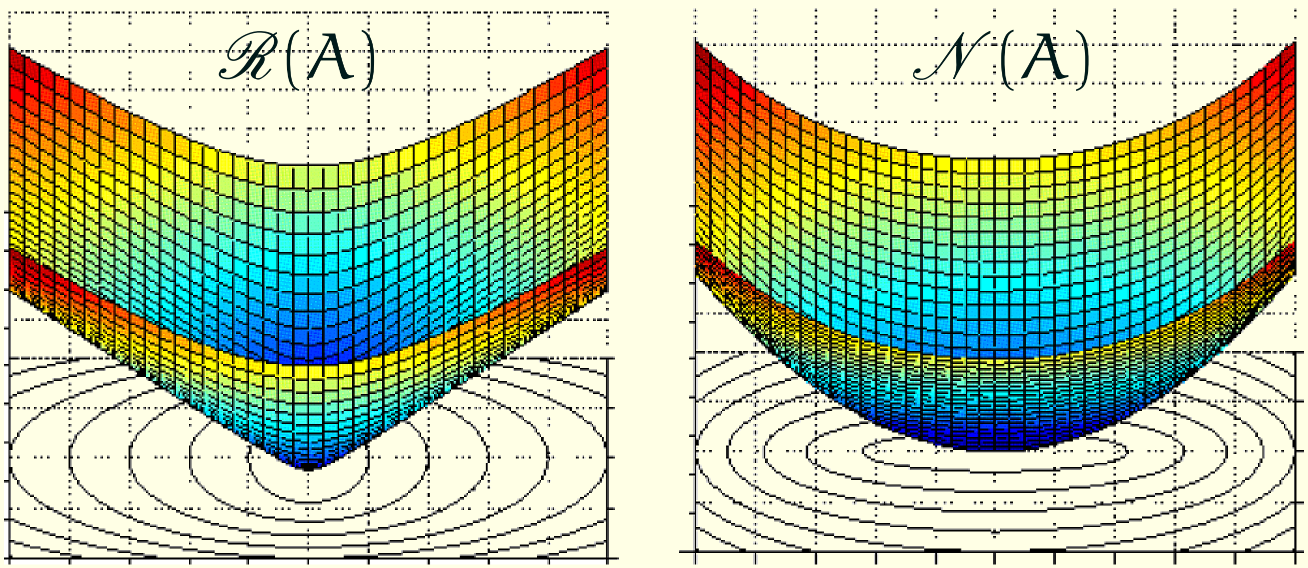
with increased stability

Extracting

Implicit Structure

VU Algorithm

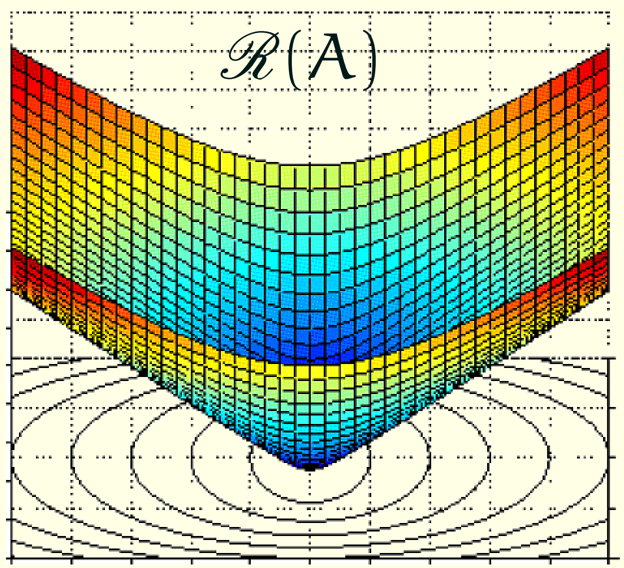
Recall that $f|_{\mathcal{N}(A)}$ is nice:



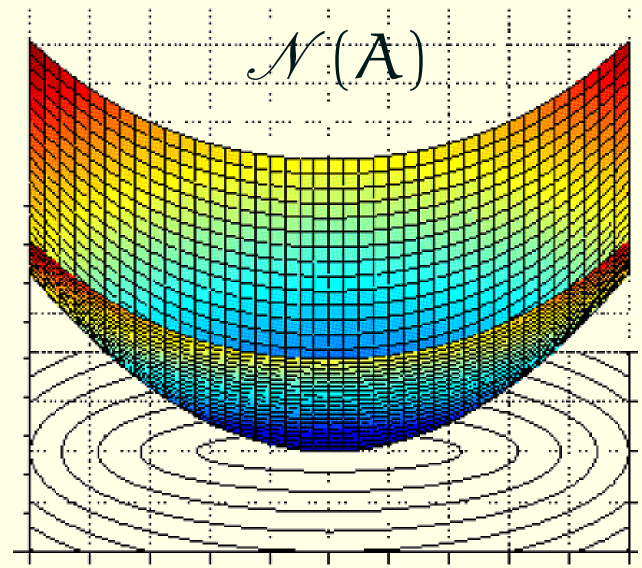
V **U**

VU Algorithm

Recall that $f|_{\mathcal{N}(A)}$ is nice:



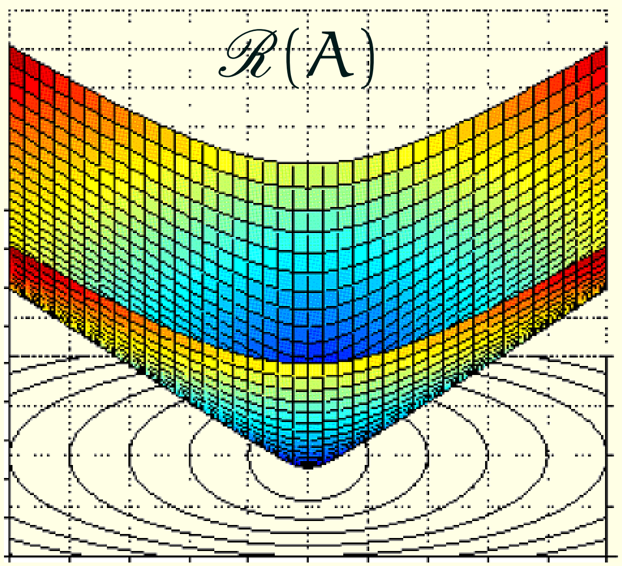
bundle QP

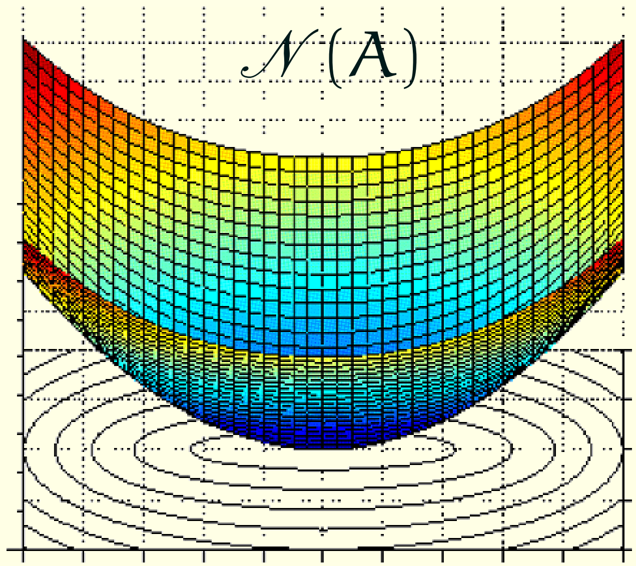


Newton-move

VU Algorithm

Recall that $f|_{\mathscr{N}(A)}$ is nice:



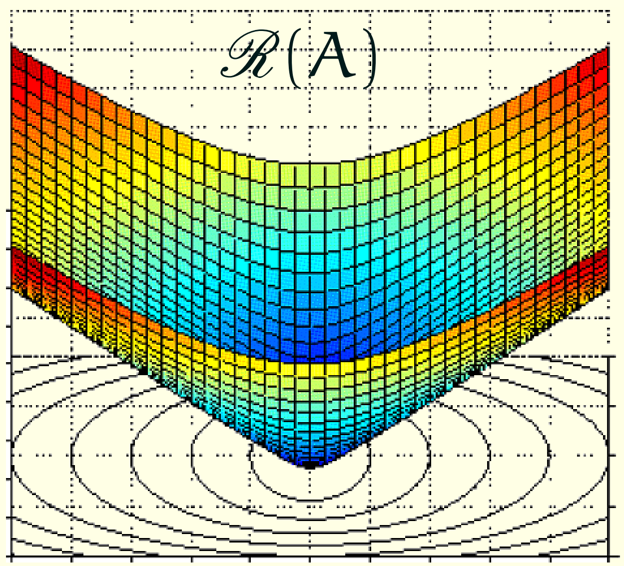


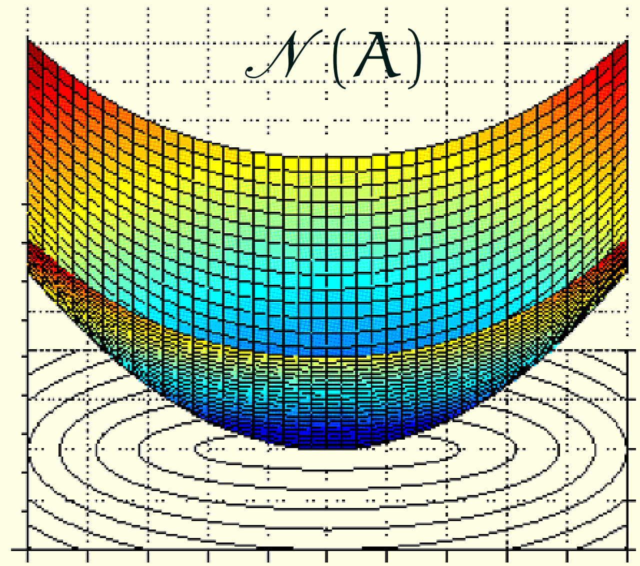
bundle QP **Newton-move**

V ?? **U**

VU Algorithm

Recall that $f|_{\mathcal{N}(A)}$ is nice:





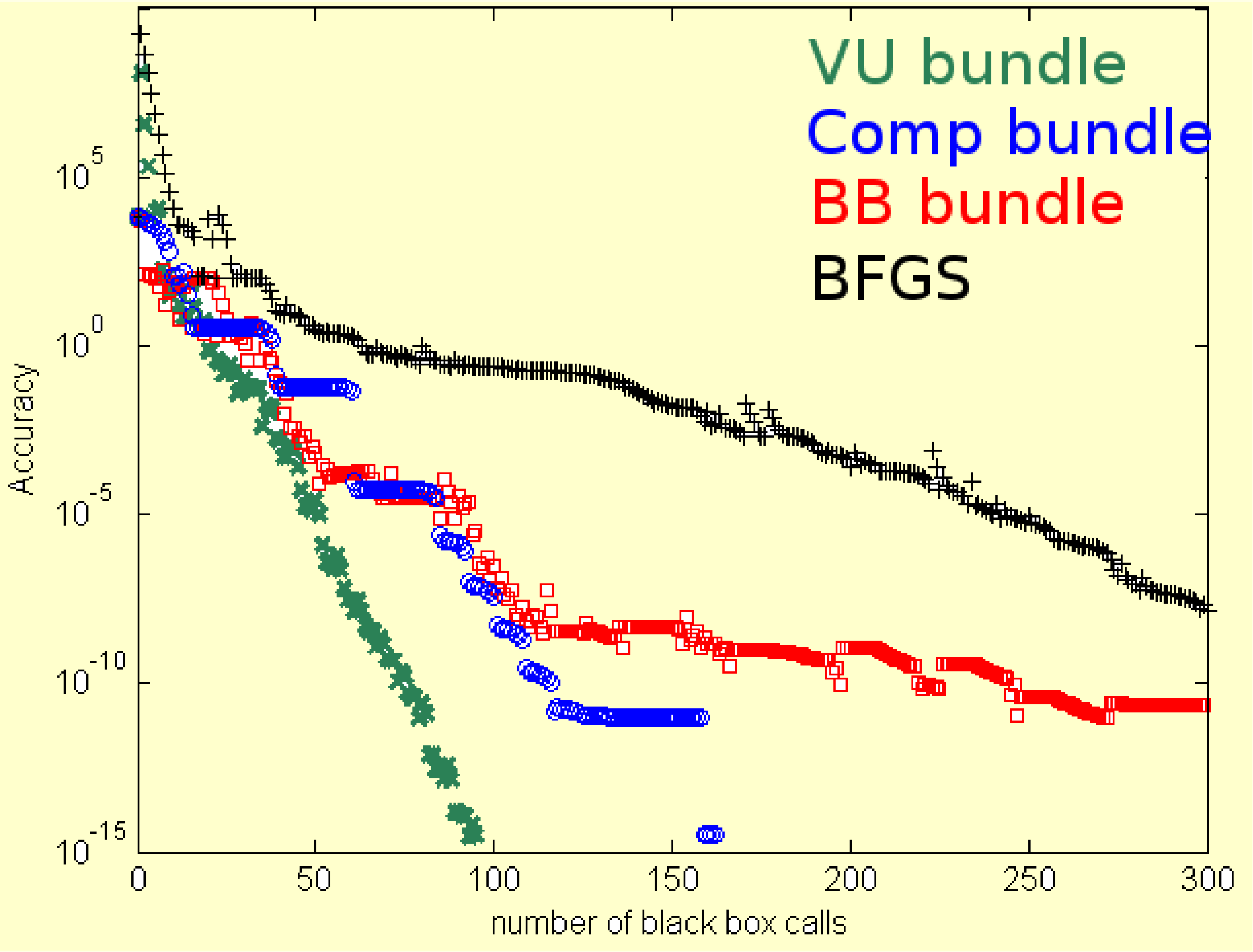
bundle QP **Newton-move**

V ?? **U**

Answer: **Bundle QP identifies the "ridge" of nonsmoothness**

Solve a 2nd QP to create a model of $\mathcal{V}$ using $\partial\varphi$

VU Algorithm: superlinear "serious" subsequence



Across borders

Constrained problems

$$\min f(x) \quad \text{s.t.} \quad c(x) \leq 0$$

φ models the Improvement Function

$$\max_x \{f(x) - f(\hat{x}), [c(x)]^+\}$$

(changes with each serious point $\hat{x}$)

Across borders

Nonconvex problems

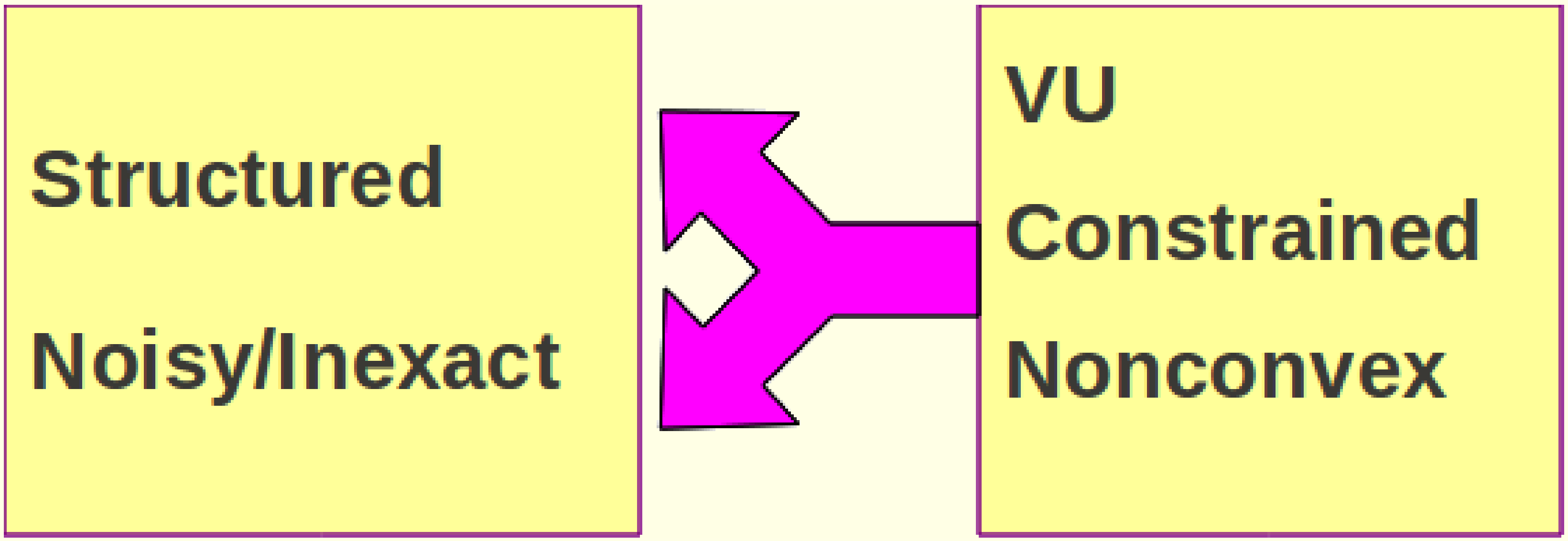
φ models the Local Convexification

$$f(x) + \frac{1}{2}\eta|x - \hat{x}|^2$$

(changes with each serious point $\hat{x}$)

Across borders

Combinations:



Closing credits: co-authors

– Robert Mifflin

– Alexandre Belloni

– Aris Daniilidis

– Grégory Emiel

– Warren Hare

– Elizabeth Karas (with A. Ribeiro)

– Claude Lemaréchal (with F. Oustry)

– Welington Oliveira (with S. Scheimberg)

– Mikhail Solodov

– Wim Van Ackooij (with R. Henrion and R. Zorgati)