Science Fiction in Nonsmooth Optimization

Robert Mifflin

http://www.math.wsu.edu/faculty/mifflin

Work with C. Sagastizábal dedicated to
Claude Lemarechal
who once said, ”superlinear convergence in nonsmooth optimization is science fiction”

SIOPT 2011, Darmstadt

Grant support: NSF DMS 0707205, AFOSR FA9550-08-1-0370 and SOARD

1 Motivation

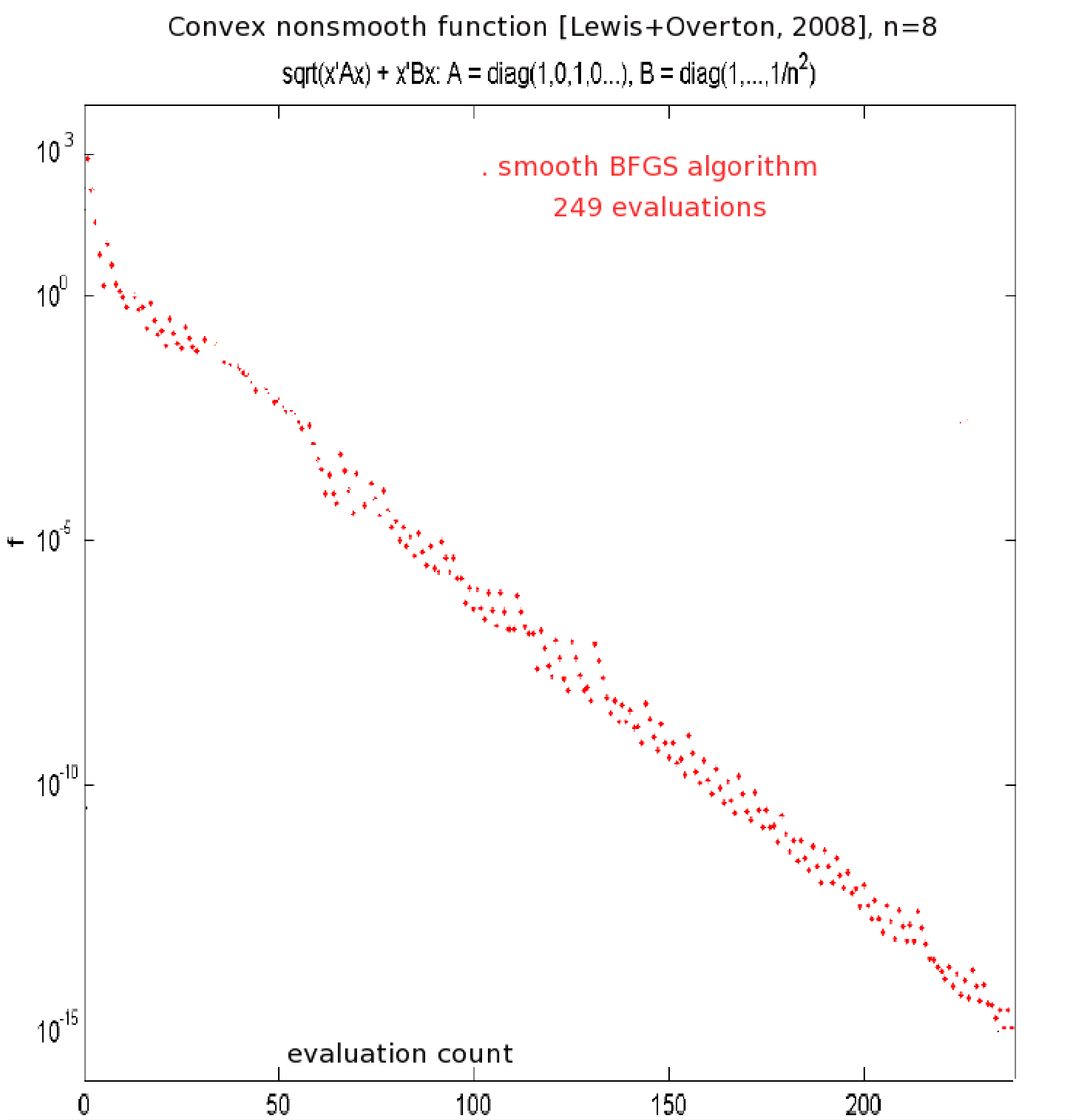
$\min_{x \in \mathbb{R}^n} f(x)$, f convex (lower C^2 in the future),

know only one subgradient of f at each x.

Fast algorithms need to identify some "curvature";

only possible if Smooth Substructure exists.

Goal is to exploit natural structure implicitly,
including nonsmoothness,
without adding extraneous structure
from barrier or smoothing functions.


Convex nonsmooth function [Lewis+Overton, 2008], n=8
sqrt(x'Ax) + x'Bx: A = diag(1,0,1,0...), B = diag(1,...,1/n^2)
. smooth BFGS algorithm
249 evaluations
f
10^3
10^0
10^-5
10^-10
10^-15
evaluation count
0
50
100
150
200

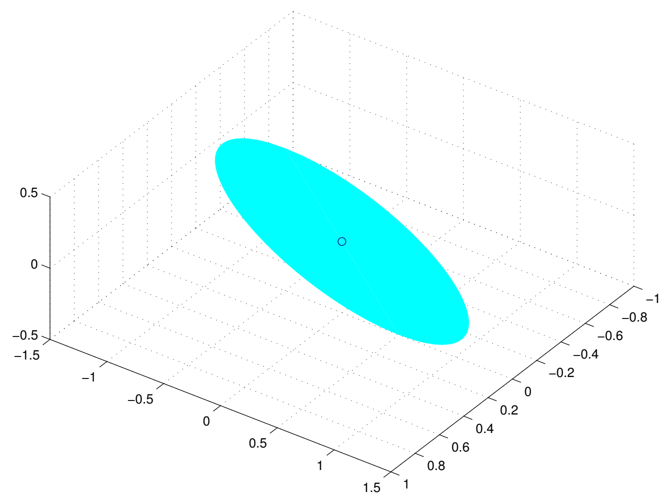
2 $\mathcal{VU}$-theory & primal-dual tracks

A pdg-structured example

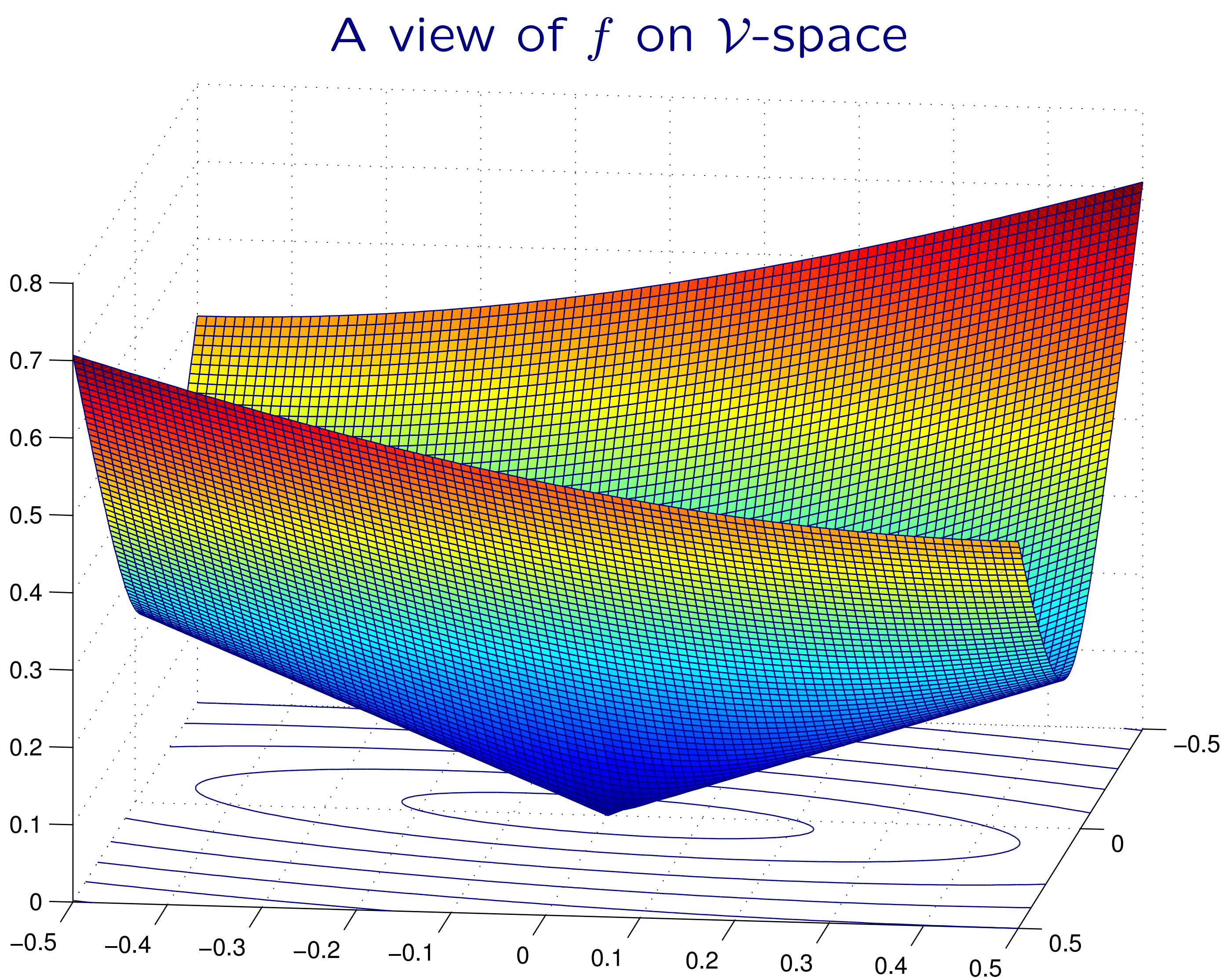
$$f(x_1, x_2, x_3) = \frac{1}{2}x_1^2 + \frac{1}{2}\sqrt{(x_1^2 - 2x_2)^2 + (x_3 - x_2)^2}$$

minimizer $\bar{x} = (0, 0, 0)$

$$\partial f((0, 0, 0))$$



For $\bar{g} \in \partial f(\bar{x})$ $\quad \mathcal{V} = lin(\partial f(\bar{x}) - \bar{g})$ **and** $\mathcal{U} := \mathcal{V}^{\perp}$



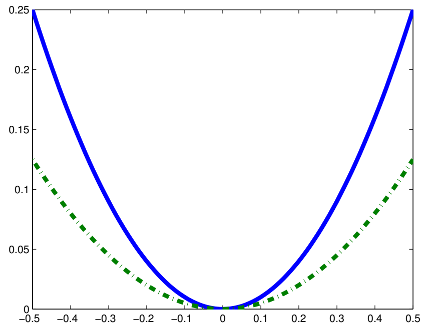
A view of f on $\mathcal{V}$-space
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
–0.5
–0.4
–0.3
–0.2
–0.1
0
0.1
0.2
0.3
0.4
0.5
–0.5
0
0.5

2.1 Primal track to $\bar{x}$

$\mathcal{U}$-Lagrangian:

$$L^{\bar{g}}_{\mathcal{U}}(u) := \inf_{v \in \mathcal{V}} \{f(\bar{x} + u \oplus v) - \langle \bar{g}, v \rangle\}$$
$$= f(\bar{x} + u \oplus v(u)) - \langle \bar{g}, v(u) \rangle$$

[Lemarechal, Oustry, Sagastizabal, 2000]



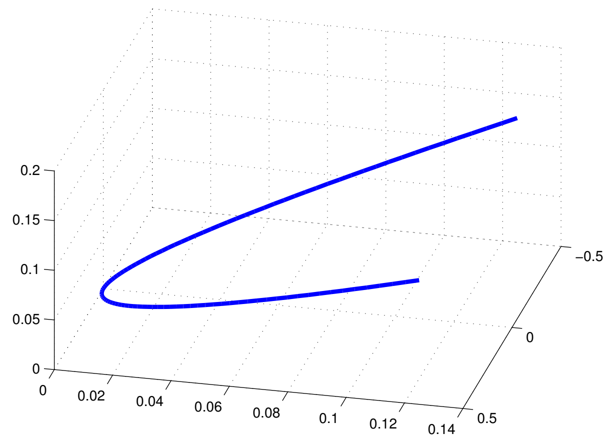
f and $L^0_{\mathcal{U}}$ on the $\mathcal{U}$-space

$\rightarrow L^{\bar{g}}_{\mathcal{U}}(0) = f(\bar{x})\,, \quad \rightarrow L^{\bar{g}}_{\mathcal{U}}(u) \in C^1(\mathcal{U})$

$\rightarrow$ minimizer $v = v(u)$ generates trajectory smooth / tangent to $\mathcal{U}$

if $\forall \bar{g} \in ri\, \partial f(\bar{x})$ the minimizer $v(u)$ is the same

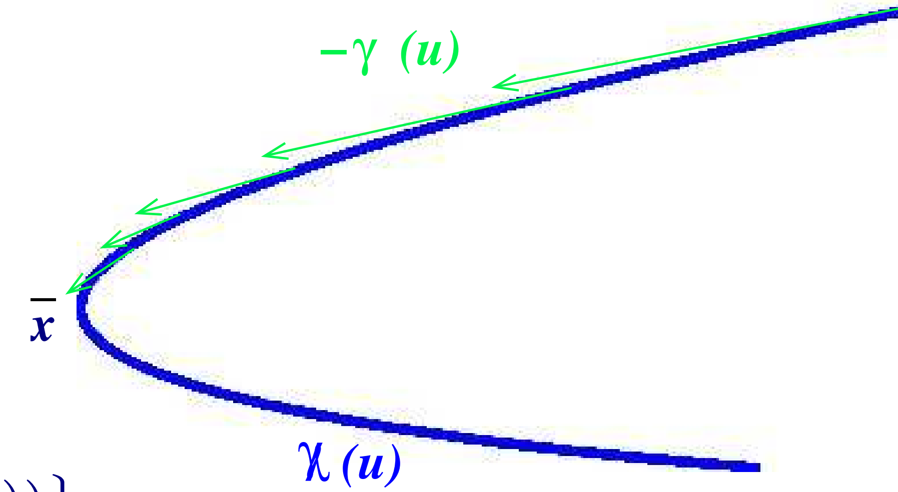
$\exists$ primal track $\chi(u) := \bar{x} + u \oplus v(u)$



with $f(\chi(u)) = L^0_{\mathcal{U}}(u)$

?what about $\nabla L^0_{\mathcal{U}}(u)$?

2.2 Dual track to $0 \in \partial f(\bar{x})$



$$\begin{aligned} \text{For } \chi(u) &= \bar{x} + u \oplus v(u) \\ \gamma(u) &:= \operatorname{argmin}\left\{|g|^2 : g \in \partial f(\chi(u))\right\} \\ \nabla L^0_{\mathcal{U}}(u) &= \mathcal{U}(\chi(u))\text{-component of } \gamma(u) \\ \big(\chi(u), \gamma(u)\big) &\to (\bar{x}, 0) \text{ as } u \to 0 \end{aligned}$$

Good primal-dual track $\leftrightarrow$ $L^0_{\mathcal{U}} \in C^2$ + $0 \in ri\, \partial f(\bar{x})$
allows for a $\chi(u)$-restricted Newton method to minimize f

3 Approximating primal-dual tracks

Fundamental theoretical result:

Proximal Points are on the primal track

If $\bar{g} = 0 \in ri\partial f(\bar{x})$, then for all $x \approx \bar{x}$ there exists $u(x)$:

$$p(x) := \operatorname{argmin}\left\{ f(y) + \frac{1}{2}\mu|y-x|^2 \right\} = \chi(u(x))$$

even with $\mu = \mu(x) : \mu(x)|x - \bar{x}| \longrightarrow 0$ as $x \longrightarrow \bar{x}$

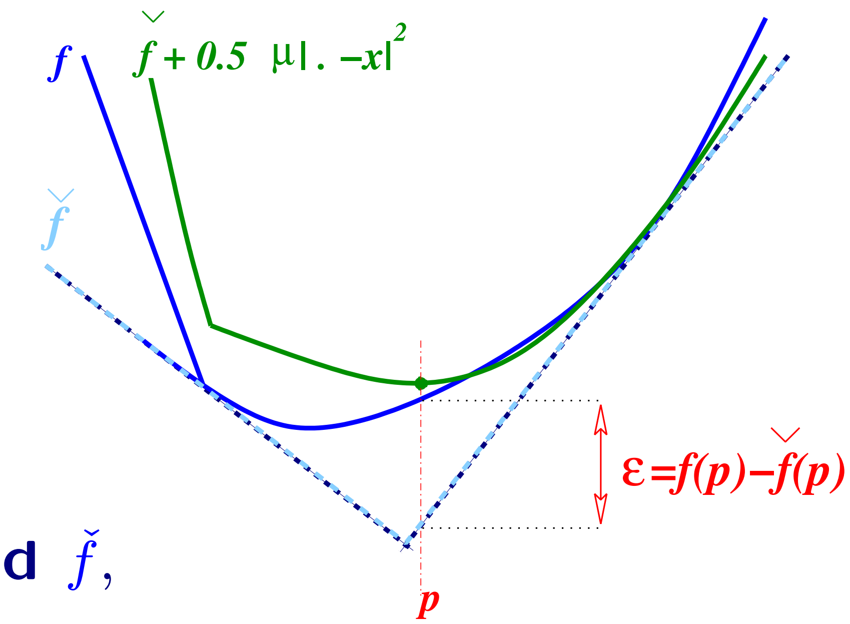
$\Rightarrow$ **use a bundle subroutine**

to approximate the prox

and estimate the pair

$(\chi(u), \gamma(u))$ **by** (p, s)

3.1 Bundle approximation



With bundle (y_i, f_i, g_i), **recursively build** $\check{f}$,
a V**-model for** f **near** x**, and find**

χ-qp **solution:** $p := argmin\left\{\check{f}(y) + \frac{1}{2}\mu|y-x|^2\right\} \approx \chi(u(x))$

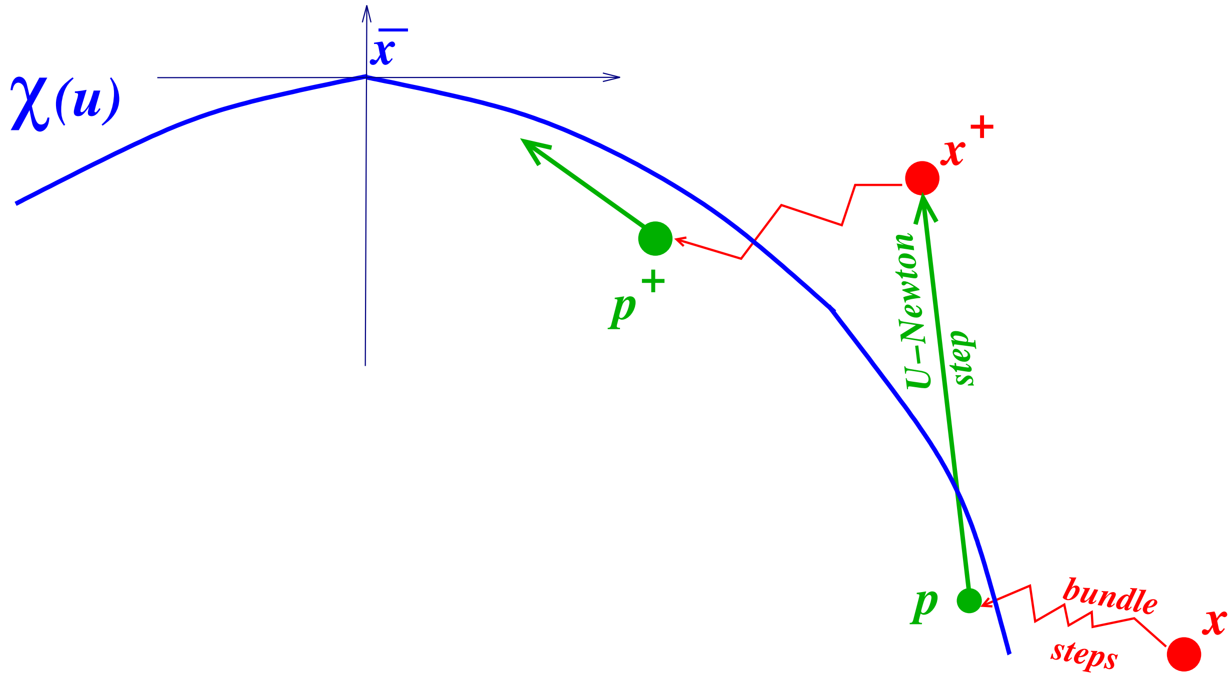
γ-qp **solution:** $s := argmin\left\{|g|^2 : g \in \partial\check{f}(p)\right\} \approx \gamma(u(x))$

UNTIL "good enough": $\varepsilon \leq (\sigma/\mu)|s|^2$

By-product: local $\mathcal{VU}$-decomposition, VU

4 Newton-like corrector-predictor $\mathcal{VU}$ algorithm

Given x and a bundle:

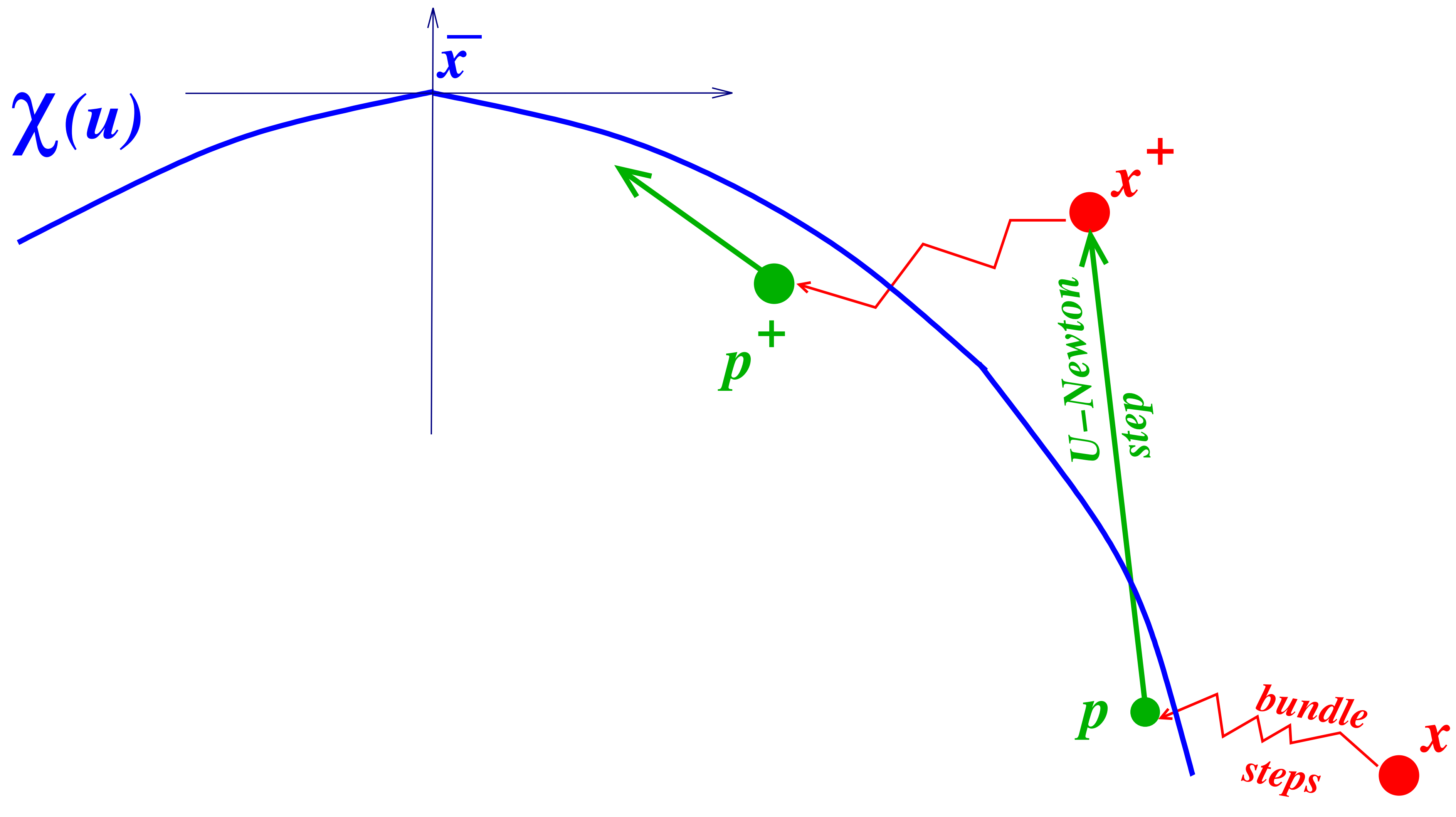


- **Corrector step:** Solve ($\chi -$ and γ-qp)'s ending

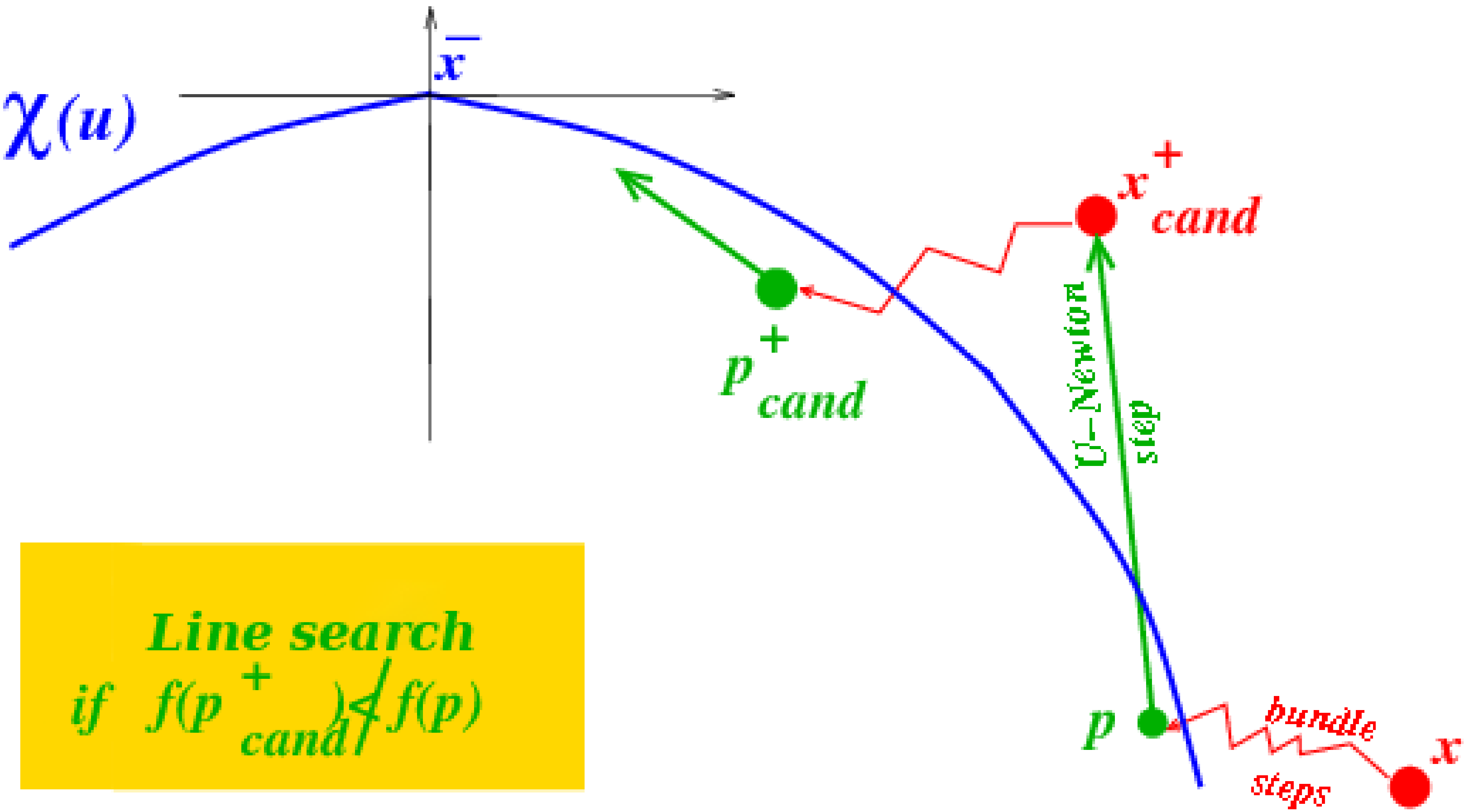
  with $p, s, \mathtt{U}$, where $\mathtt{U}^T s \approx \nabla L^0_{\mathcal{U}}$, and
  determine H, a $\mathcal{U}$-Hessian $\approx \nabla^2 L^0_{\mathcal{U}}$

- **Predictor step:** Solve $H\Delta u = -\mathtt{U}^\top s \Rightarrow x^+ = p + \mathtt{U}\Delta u$

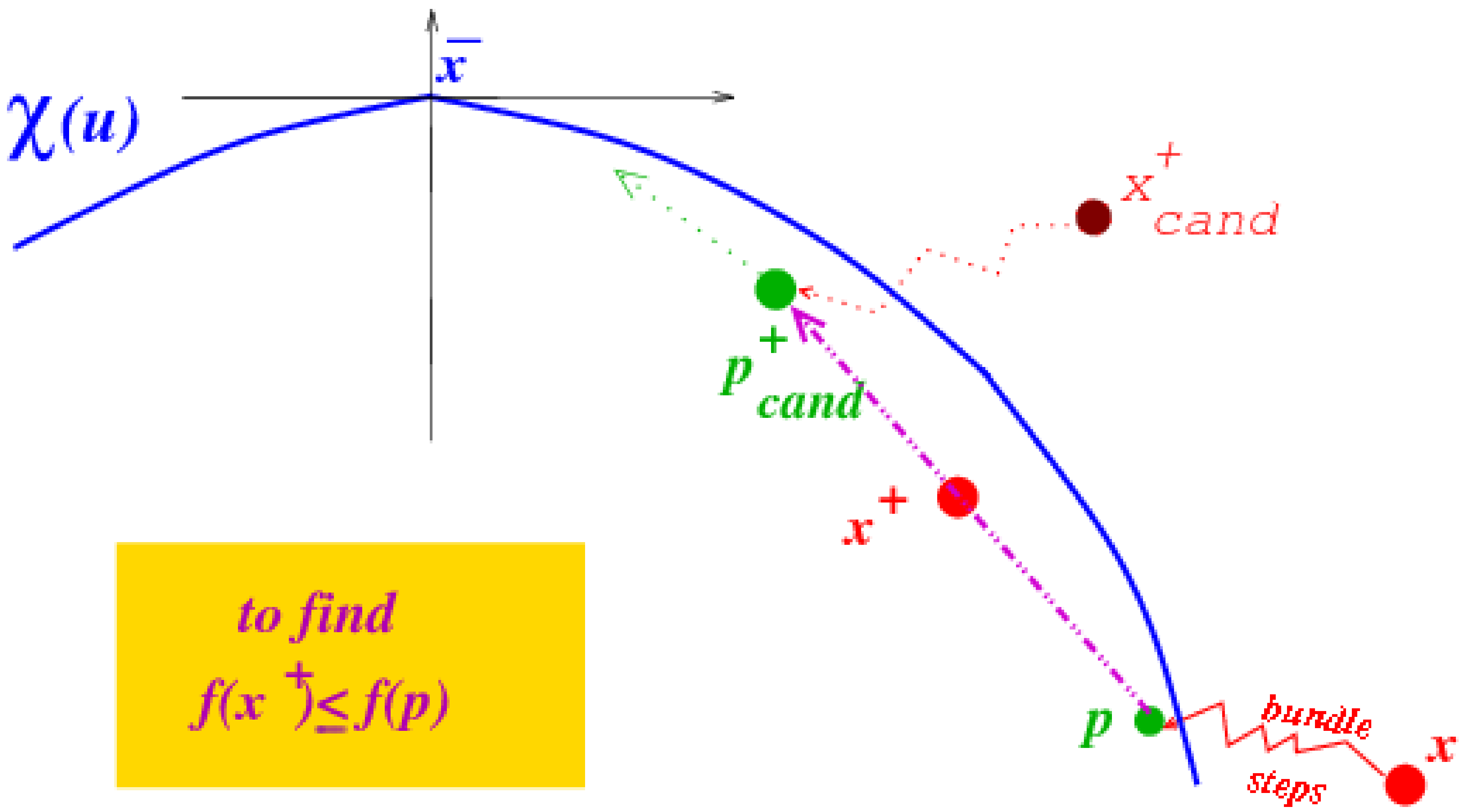
4.1 Ideal iteration



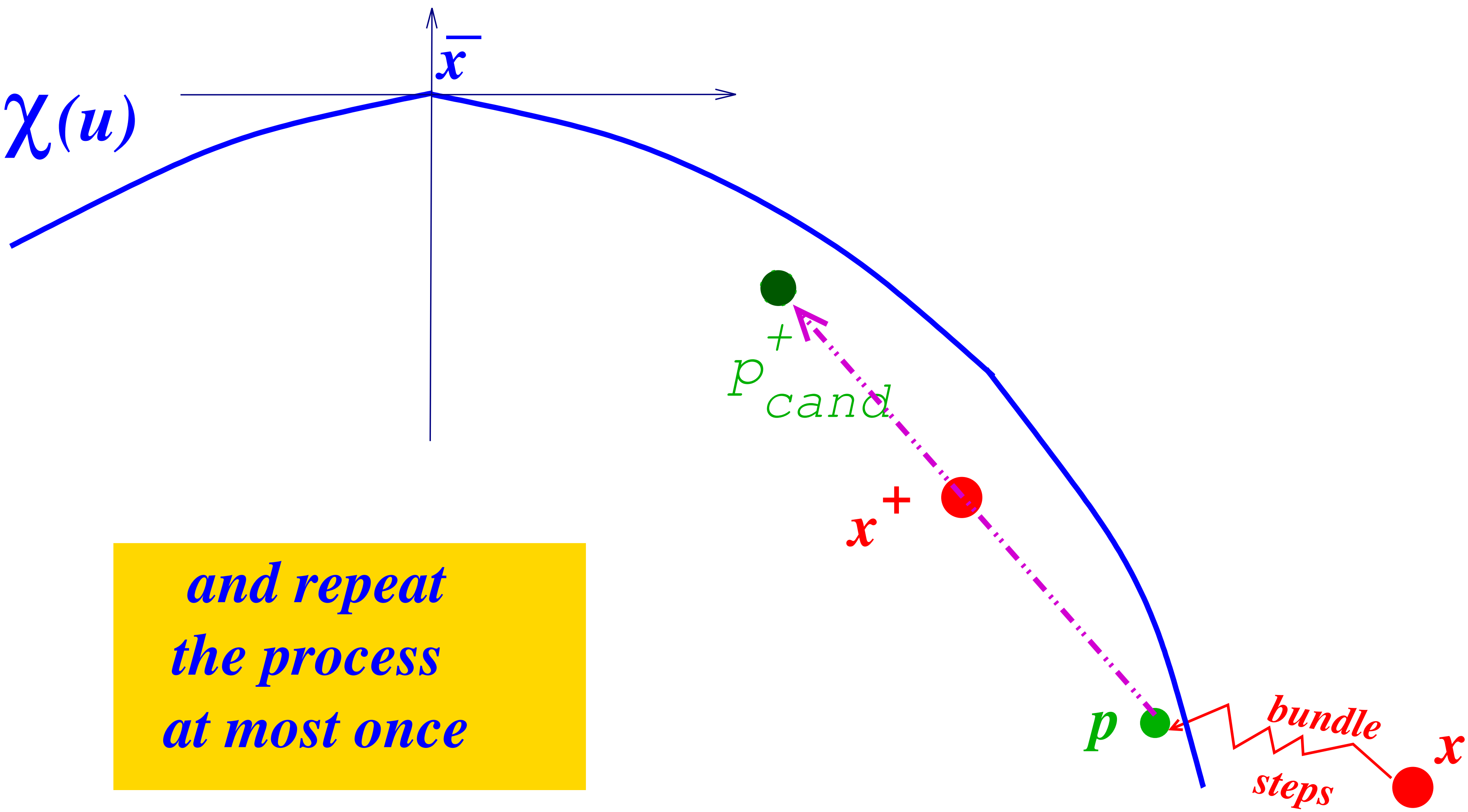
4.1 Candidates for x^+, p^+



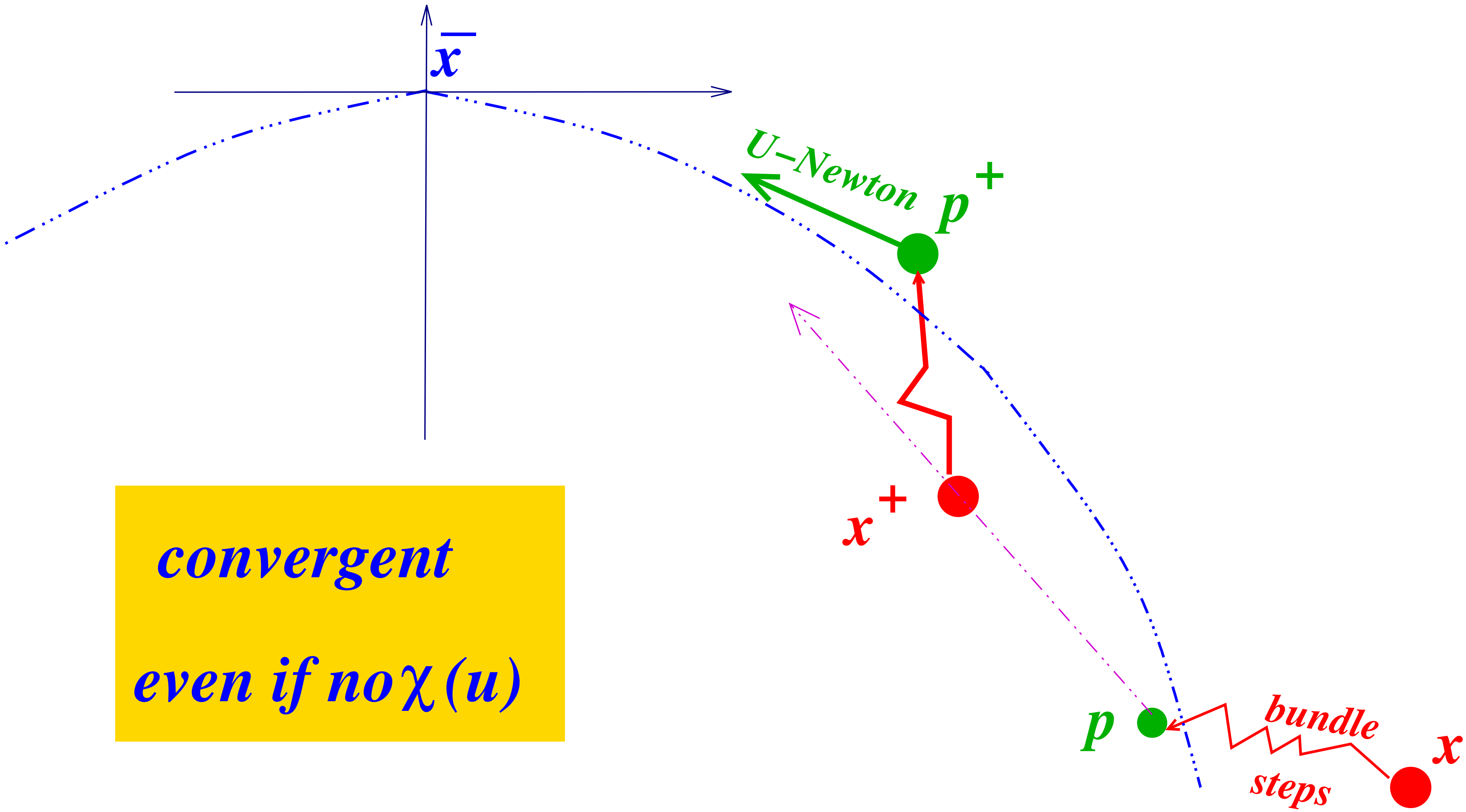
4.1 Line search if candidates fail f-descent



4.1 Line search to good x^+



4.1 Second bundle run to good p^+



4.2 Convergence properties

1. If infinite number of inner bundle steps, this sequence converges to a minimizer of f
2. If the decreasing sequence $\{f(p)\}$ is infinite, then
   - either f unbounded below,
   - or $\{s\} \to 0$ and any $acc(\{p\})$ minimizes f
3. If a primal-dual track to a strong minimizer pair $(\bar{x}, 0)$ exists and
   - $\frac{\sigma}{\mu^2} = O(|s^-|^2)$,
   - bounded $\{H^{-1}\}$,
   - $acc(\{\mathtt{U}\}) \to$ a basis for $\mathcal{U}$ ([Daniilidis, Sagastizabal, Solodov, 2009]),
   - Dennis-Moré-like condition for $\{H\}$,
   - $s - \gamma = o(|s|) \; or \; o(|\gamma|)$

then $\{p\}$ converges superlinearly to $\bar{x}$

4.3 Preliminary numerical results for a quasi-Newton version

$$H = \mathtt{U}^\top H_{qN} \mathtt{U}$$

where $H_{qN} = BFGS(p - p^-, s - s^-)$ is an $n \times n$ matrix, with updating started at or after iteration 3 when a sufficiently large curvature is found for initial scaling of the identity

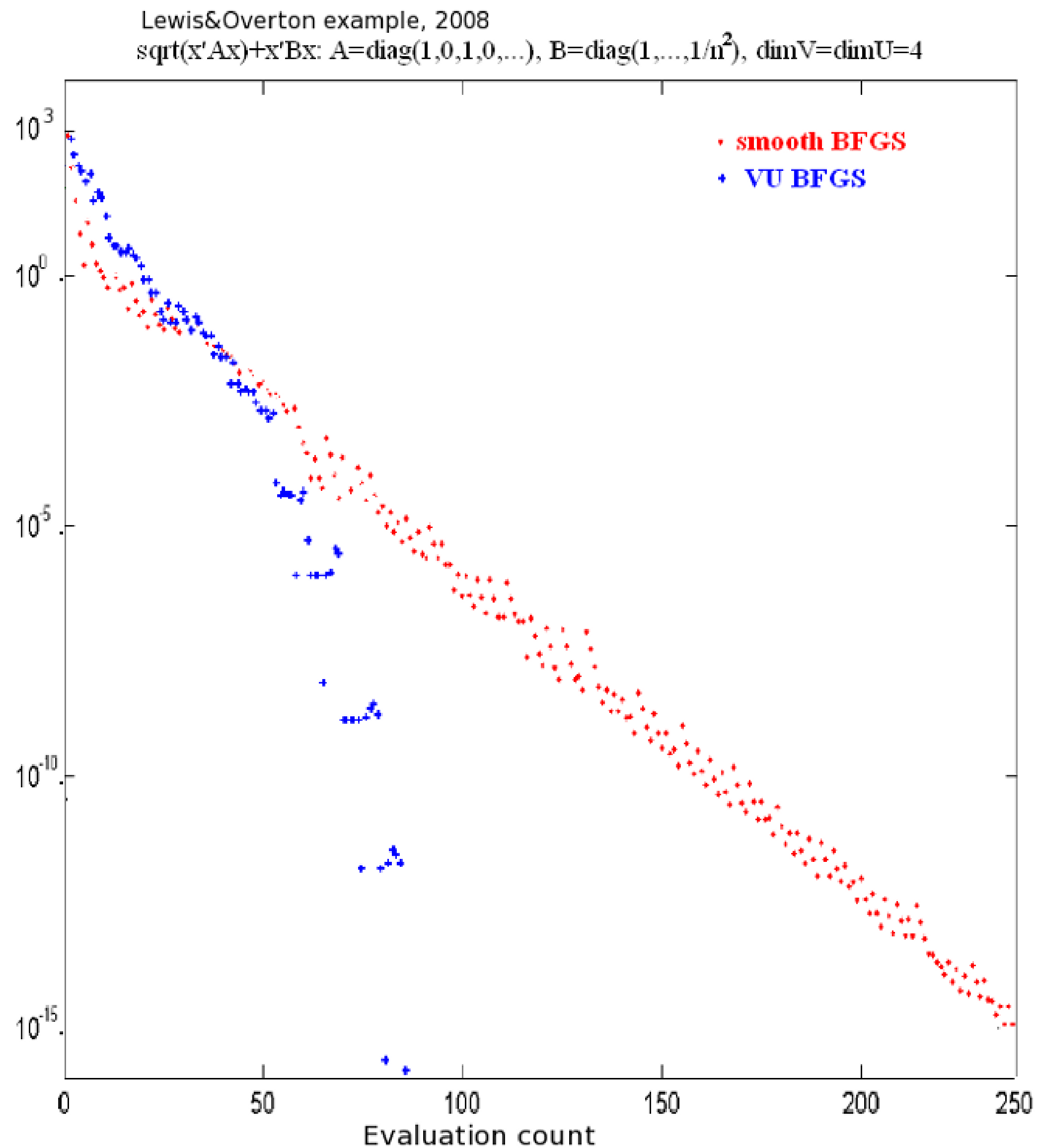
Summary of results

	2d-U1		3d-EX		3d-U2		3d-U1		3d-U0		MAXQUAD	
	f/g	Ac	f/g	Ac	f/g	Ac	f/g	Ac	f/g	Ac	f/g	Ac
N1CV2	38	7	103	7	55	7	61	7	30	7	156	8
VU.qN	19	16	35	14	36	17	27	16	32	16	72	12

N1CV2: LS, Math. Programming 76:393–410, 1997.
VU.qN: BFGS version with $\sigma = 0.5$, heuristic μ-update rules when μ too large, adequate V-approximation line search, last p replaced by best p at bundle termination and a possible line search along $\mathtt{U}\Delta u$ to satisfy a Wolfe directional derivative increase test or to not use $p + \mathtt{U}\Delta u$ as the next bundle center if f there is relatively too large

Nonsmooth Science Fiction

Smooth-BFGS 249 VU-BFGS 87



Conclusion

Important to have BOTH V and U models
and alternating V and U steps dependent on each other

Main ref.: A VU-algorithm for convex minimization, R. Mifflin and C. Sagastizábal. Math. Program. 104(2-3), pp. 583-608, 2005.

Current work: quasi-Newton, μ-adjustment and extension to semismooth functions using negative curvature estimates in appropriate V-models