Some Issues in Stochastic Variational Problems

Stephen M. Robinson

Department of Industrial and Systems Engineering
University of Wisconsin-Madison

Work supported by U. S. Air Force Research Laboratory

International Colloquium
on Stochastic Modeling and Optimization
Dedicated to Professor András Prékopa on his 80th Birthday
Rutgers Center for OR, 30 Nov - 01 Dec 2009

Outline of Topics

Overview

Here's what I hope this presentation will do:

- Explain a new and quite different way of looking at variational problems

Overview

Here's what I hope this presentation will do:

- Explain a new and quite different way of looking at variational problems
- Demonstrate some of the powerful tools that that method provides for analysis

Overview

Here's what I hope this presentation will do:

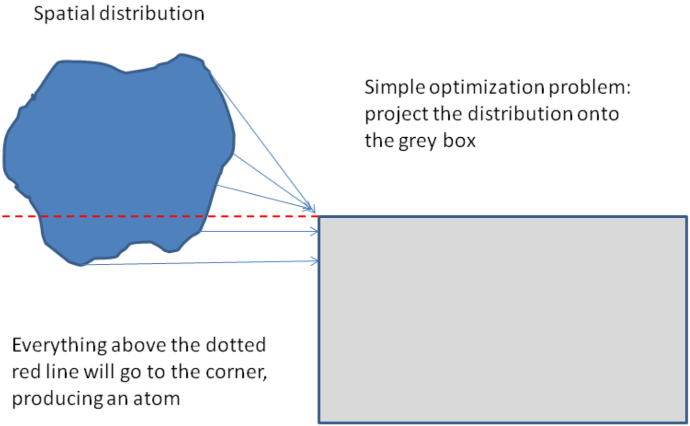
- Explain a new and quite different way of looking at variational problems
- Demonstrate some of the powerful tools that that method provides for analysis
- At the end, suggest how these tools can be used for analysis of stochastic variational problems

Overview

Here's what I hope this presentation will do:

- Explain a new and quite different way of looking at variational problems
- Demonstrate some of the powerful tools that that method provides for analysis
- At the end, suggest how these tools can be used for analysis of stochastic variational problems
- And encourage those who know more than I do to explore some of those possibilities

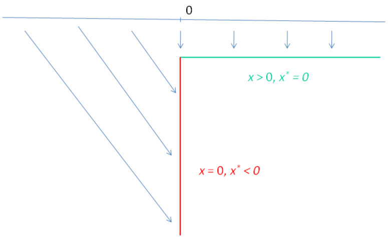
Example: Projecting a Cloud



(King & Rockafellar, 1993)

Simpler Example: What's Happening?

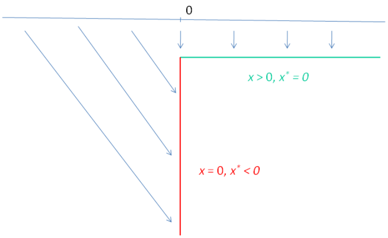
Projecting from the real line to a halfline



- Our problem: $\min\{(1/2)|x - r|^2 \mid x \in \mathbb{R}_+\}$

Simpler Example: What's Happening?

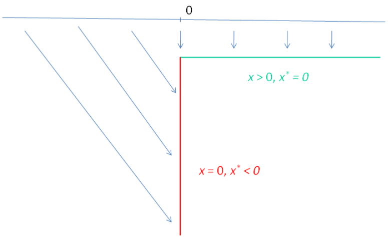
Projecting from the real line to a halfline



- Our problem: $\min\{(1/2)|x - r|^2 \mid x \in \mathbb{R}_+\}$
- Optimality condition: $0 \in x - r + N_{\mathbb{R}_+}(x)$

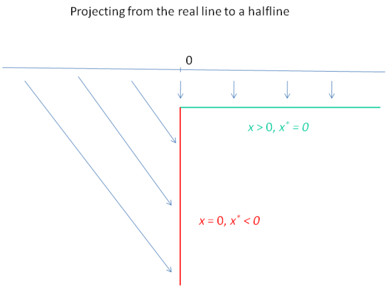
Simpler Example: What's Happening?

Projecting from the real line to a halfline



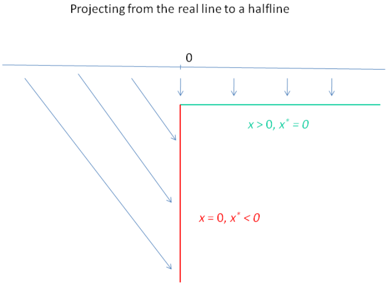
- Our problem: $\min\{(1/2)|x - r|^2 \mid x \in \mathbb{R}_+\}$
- Optimality condition: $0 \in x - r + N_{\mathbb{R}_+}(x)$
- Rewrite optimality condition: $r = x + x^*, \ (x, x^*) \in \text{gph}\, N_{\mathbb{R}_+}(x)$

Simpler Example: What's Happening?



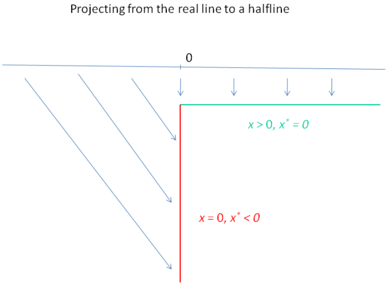
- Our problem: $\min\{(1/2)|x-r|^2 \mid x \in \mathbb{R}_+\}$
- Optimality condition: $0 \in x - r + N_{\mathbb{R}_+}(x)$
- Rewrite optimality condition: $r = x + x^*,\ (x, x^*) \in \operatorname{gph} N_{\mathbb{R}_+}(x)$
- Figure shows x and x^*

Simpler Example: What's Happening?



- Our problem: $\min\{(1/2)|x-r|^2 \mid x \in \mathbb{R}_+\}$
- Optimality condition: $0 \in x - r + N_{\mathbb{R}_+}(x)$
- Rewrite optimality condition: $r = x + x^*,\ (x, x^*) \in \text{gph } N_{\mathbb{R}_+}(x)$
- Figure shows x and x^*
- When we only look at x we're *throwing away half the problem*

Simpler Example: What's Happening?



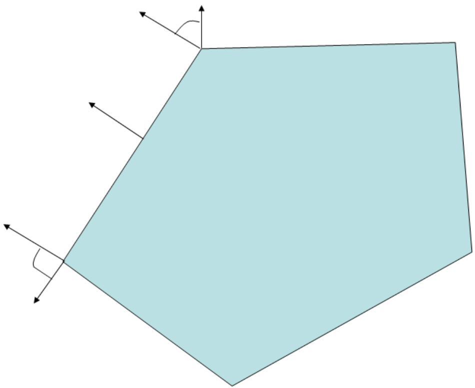
- Our problem: $\min\{(1/2)|x - r|^2 \mid x \in \mathbb{R}_+\}$
- Optimality condition: $0 \in x - r + N_{\mathbb{R}_+}(x)$
- Rewrite optimality condition: $r = x + x^*,\ (x, x^*) \in \text{gph}\, N_{\mathbb{R}_+}(x)$
- Figure shows x and x^*
- When we only look at x we're *throwing away half the problem*
- That's what caused our difficulty with the cloud

Some Perspective

- First, this projection problem is part of a general class of *variational conditions.*
- In this case, it's a *variational inequality:* we have a closed convex set C (here, $\mathbb{R}^1_+$) and a function $f : \mathbb{R}^1 \to \mathbb{R}^1$ (here, $f(x) = x - r$), and we want to find a point x_* such that $-f(x_*)$ is (outwardly) normal to C.
- A convenient way to express this: find a solution of $0 \in f(x) + N_C(x)$, where

$$N_C(x) = \begin{cases} \{x^* \mid \text{For each } c \in C,\ \langle c - x, x^* \rangle \leq 0\} & \text{if } x \in C, \\ \emptyset & \text{if } x \notin C. \end{cases}$$

Normal Cones Illustrated



Equations on Graphs

- This expression $0 \in f(x) + N_C(x)$ looks a lot like an equation, except for the multivalued normal-cone operator N_C

Equations on Graphs

- This expression $0 \in f(x) + N_C(x)$ looks a lot like an equation, except for the multivalued normal-cone operator N_C
- That's easy to fix: write

$$f(x) + x^* = 0, \qquad (x, x^*) \in \operatorname{gph} N_C,$$

where $\operatorname{gph} N_C = \{(x, x^*) \mid x^* \in N_C(x)$

Equations on Graphs

- This expression $0 \in f(x) + N_C(x)$ looks a lot like an equation, except for the multivalued normal-cone operator N_C
- That's easy to fix: write

$$f(x) + x^* = 0, \qquad (x, x^*) \in \operatorname{gph} N_C,$$

  where $\operatorname{gph} N_C = \{(x, x^*) \mid x^* \in N_C(x)\}$
- Now we have a real equation, but the underlying set is $\operatorname{gph} N_C$ instead of some space $\mathbb{R}^k$

Equations on Graphs

- This expression $0 \in f(x) + N_C(x)$ looks a lot like an equation, except for the multivalued normal-cone operator N_C
- That's easy to fix: write

$$f(x) + x^* = 0, \qquad (x, x^*) \in \operatorname{gph} N_C,$$

  where $\operatorname{gph} N_C = \{(x, x^*) \mid x^* \in N_C(x)\}$
- Now we have a real equation, but the underlying set is $\operatorname{gph} N_C$ instead of some space $\mathbb{R}^k$
- We would expect this trivial reformulation to be worthwhile *only* if $\operatorname{gph} N_C$ had hidden structure that we could somehow exploit

Equations on Graphs

- This expression $0 \in f(x) + N_C(x)$ looks a lot like an equation, except for the multivalued normal-cone operator N_C
- That's easy to fix: write
$$f(x) + x^* = 0, \qquad (x, x^*) \in \operatorname{gph} N_C,$$
where $\operatorname{gph} N_C = \{(x, x^*) \mid x^* \in N_C(x)\}$
- Now we have a real equation, but the underlying set is $\operatorname{gph} N_C$ instead of some space $\mathbb{R}^k$
- We would expect this trivial reformulation to be worthwhile *only* if $\operatorname{gph} N_C$ had hidden structure that we could somehow exploit
- It does: in fact, it has a very rich structure

This formulation includes a wide class of problems

- First-order optimality conditions for nonlinear programming problems (with multipliers, if needed, to accommodate nonlinear constraints)
- Linear and nonlinear complementarity problems
- Traffic equilibrium problems
- Stationarity conditions for other Nash equilibrium problems, including those from some games
- Equilibrium problems from computational economics

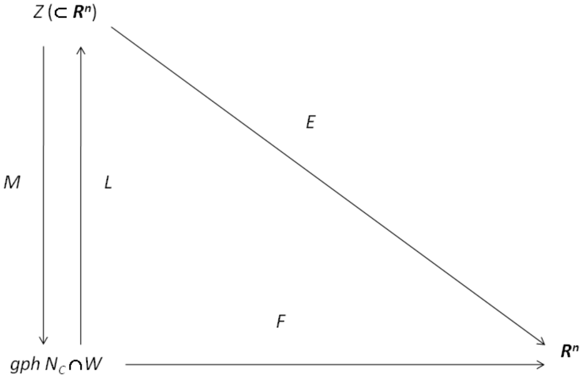
The first difficulty

- Again:

$$F(x, x^*) = f(x) + x^* = 0, \qquad (x, x^*) \in \operatorname{gph} N_C$$

- We seem to have too many variables to solve for (x, x^*): we're sending $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$
- But there's another constraint: $(x, x^*) \in \operatorname{gph} N_C$
- We have to combine these in order to do the analysis
- Next slide gives a picture of the combination

Lifting F to a nonsmooth function E



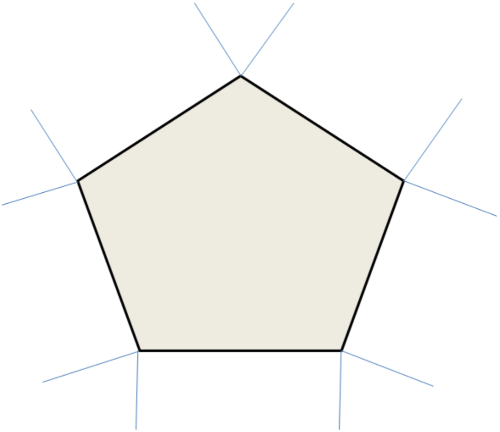
Finding Lipschitz Homeomorphisms L and M

- For $z \in \mathbb{R}^n$ let $\Pi_C(z)$ be the Euclidean projector on C, and define

$$M(z) = [\Pi_C(z), I - \Pi_C(z)], \qquad L(x, x^*) = x + x^*$$

- Minty's theorem says M is a Lipschitz homeomorphism of $\mathbb{R}^n$ onto gph N_C, with inverse L
- When C is polyhedral convex, the map M is *piecewise affine*
- The subsets on which M is affine form a polyhedral subdivision of $\mathbb{R}^n$ called the *normal manifold*

The normal manifold of a pentagon



Summary of the Formulation

- Start with the given function F and set C; fix $w_0 = (x_0, x_0^*) \in \operatorname{gph} N_C$ with $F(x_0, x_0^*) = 0$
- Construct the Lipschitz homeomorphisms L and M
- Construct $E = (F \circ M) : Z \to \mathbb{R}^n$, which is the map we will analyze
- As M will usually be nonsmooth, so will be E, even if F is smooth. This is the price we pay for dealing with nasty graphs

Well-behaved equations

- In ordinary analysis, a nice function f from an open subset of $\mathbb{R}^n$ into $\mathbb{R}^n$ could be a (local) homeomorphism: f and f^{-1} are both (locally) single-valued and continuous
- Even better, it could be a Lipschitz homeomorphism: f and f^{-1} each obey a Lipschitz condition
- With C^1 functions from $\mathbb{R}^n$ to $\mathbb{R}^n$, the inverse function theorem says we have a local Lipschitz homeomorphism at x_0 when the derivative $df(x_0)$ is nonsingular
- This theorem is the foundation for local analysis of C^1 functions, with innumerable applications

The situation with variational problems

- In nonlinear programming, complementarity, or other equilibrium problems we typically *do not* have anything like this, even with very nice problems (e.g., the cloud, or the projection problem in $\mathbb{R}^1$)

The situation with variational problems

- In nonlinear programming, complementarity, or other equilibrium problems we typically *do not* have anything like this, even with very nice problems (e.g., the cloud, or the projection problem in $\mathbb{R}^1$)
- But with this formulation, *we do,* because we include *both* x and its conjugate variable x^*

The situation with variational problems

- In nonlinear programming, complementarity, or other equilibrium problems we typically *do not* have anything like this, even with very nice problems (e.g., the cloud, or the projection problem in $\mathbb{R}^1$)
- But with this formulation, *we do,* because we include *both* x and its conjugate variable x^*
- Next slide explains conditions

The situation with variational problems

- In nonlinear programming, complementarity, or other equilibrium problems we typically *do not* have anything like this, even with very nice problems (e.g., the cloud, or the projection problem in $\mathbb{R}^1$)
- But with this formulation, *we do,* because we include *both* x and its conjugate variable x^*
- Next slide explains conditions
- This is a strong argument for looking at variational problems in this way, rather than in the traditional way

Nonsingularity for a piecewise affine function

- When is a piecewise affine function from a normal manifold $\mathcal{N}_C$ for a polyhedral convex $C \subset \mathbb{R}^n$ to $\mathbb{R}^n$ a Lipschitz homeomorphism?
- On each n-cell of the manifold, the function has an affine representative; the linear part of that affine function has a determinant
- f is a Lipschitz homeomorphism *if and only if* those determinants all have the same nonzero sign (so that f is *coherently oriented*)
- This extends the classical case, in which there is just one n-cell ($\mathbb{R}^n$)

What about nonlinear problems?

- Just as in the classical case, a problem with a C^1 function is a local Lipschitz homeomorphism if and only if the linearized problem has that property
- We linearize $f(x) + x^*$ by passing to the problem $f(x_0) + df(x_0)(x - x_0) + x^*$
- The proof of this nonlinear result is a little harder than the proof for the classical case
- But the proof of the coherent orientation test for piecewise affine problems is *very much* harder than that for the classical (linear) case

This gives us a good set of tools

- With usable inverse and implicit function theorems, we can
  - Do convergence analysis for algorithms,
  - Perform sensitivity analysis,
  - Formulate methods for time-dependent problems,
  - And many other things

This gives us a good set of tools

- With usable inverse and implicit function theorems, we can
  - Do convergence analysis for algorithms,
  - Perform sensitivity analysis,
  - Formulate methods for time-dependent problems,
  - And many other things
- And, the analysis is generally very much like that for classical problems, though technically harder: we mostly use single-valued functions, and we can use the extensive knowledge that is already in place for such problems

This gives us a good set of tools

- With usable inverse and implicit function theorems, we can
  - Do convergence analysis for algorithms,
  - Perform sensitivity analysis,
  - Formulate methods for time-dependent problems,
  - And many other things
- And, the analysis is generally very much like that for classical problems, though technically harder: we mostly use single-valued functions, and we can use the extensive knowledge that is already in place for such problems
- In the final section we'll look at some possible applications to stochastic problems

A stochastic variational problem

- Suppose we have a vector-valued stochastic process $\{f_n(\omega, x) \in \mathbb{R}^m \mid n = 1, 2, \ldots\}$ with the following properties
- For all $n \geq 1$ and $x \in \mathbb{R}^k$, the random variables $f_n(\omega, x)$ are defined on a common probability space $(\Omega, \mathcal{F}, P)$, and for almost all ω the $f_n(\omega, \cdot\,)$ converge pointwise to a deterministic function $f(\cdot)$
- We look for a point x_0 such that the function f satisfies $0 \in f(x) + N_C(x)$, where C is polyhedral convex
- Motivation: the f_n are estimates obtained by simulation

One approach to solution (Gürkan *et al.*, 1999)

- Fix a large n and a sample point ω
- Solve the deterministic variational inequality with $f(\cdot) = f_n(\omega, \cdot)$
- Take the solution $x_n(\omega)$ as an estimate of x_0
- One can give conditions ensuring that with probability 1, when n is sufficiently large the $x_n(\omega)$ exist and are close to x_0
- This approach has been applied to energy market problems (interruptions in natural gas supply), as well as option pricing and network design, among other areas
- Early versions of some of the results already discussed provided the justification for that analysis

Other examples

- One can use the theory described here in constructing confidence regions for variational problems (Demir, 2000)
- A slightly more comprehensive form provides tools for analyzing the behavior of robust statistical estimators
- There are many other possibilities

The point: in the hands of people more expert than I am, these tools could extend our ability to analyze stochastic problems where variational behavior is a key aspect.