Calmness in stochastic programming – exact penalization and sample approximation techniques

Martin Branda

Charles University in Prague
Faculty of Mathematics and Physics
Department of Probability and Mathematical Statistics

Stochastic Programming and Approximation
21 February 2013, Prague
1 Exact penalty method and calmness

2 Stochastic programming formulations

3 Relations between formulations

4 Sample approximations using Monte-Carlo techniques

5 References
Contents

1 Exact penalty method and calmness

2 Stochastic programming formulations

3 Relations between formulations

4 Sample approximations using Monte-Carlo techniques

5 References
Basis of the exact penalty method

Nonlinear programming problem

\[
\min_{x \in X} f(x) \\
\text{s.t.} \\
g_j(x) \leq 0, j = 1, \ldots, m,
\]

where \( f, g_j : \mathbb{R}^n \to \mathbb{R}, \ X \subseteq \mathbb{R}^n \). Corresponding penalty problem

\[
\min_{x \in X} f(x) + N \cdot \alpha(x),
\]

where

\[
\alpha(x) = \sum_{j=1}^{m} |g_j(x)|_+^p, \quad p \in \mathbb{N}.
\]
Nonlinear programming problem

\[
\min_{x \in X} f(x)
\]

s.t.

\[
g_j(x) \leq 0, \ j = 1, \ldots, m,
\]

where \( f, g_j : \mathbb{R}^n \to \mathbb{R}, \ X \subseteq \mathbb{R}^n \). Corresponding penalty problem

\[
\min_{x \in X} f(x) + N \cdot \alpha(x),
\]

where

\[
\alpha(x) = \sum_{j=1}^{m} |g_j(x)|^p_+, \ p \in \mathbb{N}.
\]
Exact penalty method and calmness

Exterior penalty method

Bazaraa et al. (2006), Theorem 9.2.2 (also for equality constraints):

Proposition

Let \( f, g_j \) be continuous, \( X \neq \emptyset \), the underlying problem have a feasible solution. Assume that for each \( N \) there is a solution \( x_N \in X \) of the penalty problem and \( \{x_N\} \) is contained in a compact subset of \( X \). Then

\[
\min_{x \in X} \{f(x) : g_j(x) \leq 0, j = 1, \ldots, m\} = \sup_{N \geq 0} \theta(N) = \lim_{N \to \infty} \theta(N),
\]

where

\[
\theta(N) = \min_{x \in X} f(x) + N \cdot \alpha(x).
\]

Furthermore, the limit of any convergent subsequence of \( \{x_N\} \) is an optimal solution to the original problem, and \( N \cdot \alpha(x_N) \to 0 \) as \( N \to \infty \).

Remark

If for some \( N > 0 \) it holds \( \alpha(x_N) = 0 \), then \( x_N \) is the optimal solution of the NLP, see Bazaraa et al. (2006), Corollary 9.2.2. It is a question, how to ensure this situations known as “exact penalization” in general.

Bazaraa et al. (2006), Theorem 9.3.1 (also for equality constraints):

Proposition

Let \((x^*, v^*) \in \mathbb{R}^n \times \mathbb{R}^m_+\) be a KKT point. Moreover, suppose that \( f, g_j \) are convex functions. Then for \( N \geq \max_j v^*_j \), \( x^* \) minimizes also the penalized objective with \( p = 1 \) (L1 penalty).
Remark

If for some $N > 0$ it holds $\alpha(x_N) = 0$, then $x_N$ is the optimal solution of the NLP, see Bazaraa et al. (2006), Corollary 9.2.2. It is a question, how to ensure this situations known as “exact penalization” in general.

Bazaraa et al. (2006), Theorem 9.3.1 (also for equality constraints):

Proposition

Let $(x^*, v^*) \in \mathbb{R}^n \times \mathbb{R}_+^m$ be a KKT point. Moreover, suppose that $f, g_j$ are convex functions. Then for $N \geq \max_j v_j^*$, $x^*$ minimizes also the penalized objective with $p = 1$ ($L_1$ penalty).
Calmness

- Calm problems
- Calm set-valued mappings (graphs)
- Calm functions
A general relaxed mathematical program

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad F(x) + u \in \Lambda
\end{align*}
\]

\(f : \mathbb{R}^n \to \mathbb{R}, \ F : \mathbb{R}^n \to \mathbb{R}^m, \ \Lambda \subseteq \mathbb{R}^m \) closed, \( u \in \mathbb{R}^m \). Underlying problem for \( u = 0 \).

We denote \( d_\Lambda (x) = \min_{x' \in \Lambda} \| x - x' \| \).
Burke (1991a), Definition 1.1:

**Definition**

Let $x^*$ be feasible for the unperturbed problem. Then the problem is said to be **calm** at $x^*$ if there exist constant $\tilde{N} \geq 0$ (modulus) and $\epsilon > 0$ (radius) such that for all $(x, u) \in \mathbb{R}^n \times \mathbb{R}^m$ satisfying $x \in B_\epsilon(x^*)$ and $F(x) + u \in \Lambda$, one has

$$f(x) + \tilde{N} \|u\| \geq f(x^*).$$

Note that then $x^*$ is necessarily a local solution to the unperturbed problem.
Exact penalization

Burke (1991a), Theorem 1.1:

Proposition

Let $x^*$ be feasible for the unperturbed problem, i.e. with $u = 0$. Then the unperturbed problem is calm at $x^*$ with modulus $\tilde{N}$ and radius $\epsilon > 0$ if and only if $x^*$ is a local minimum of the function

$$f(x) + N \cdot d_{\Lambda}(F(x))$$

over the neighbourhood $B_\epsilon(x^*)$ for all $N \geq \tilde{N}$.

Penalty function $d_{\Lambda}(x) = \min_{x' \in \Lambda} \|x - x'\|$. 
Lipschitz-like properties of set-valued mappings

Set valued mapping (multifunction) $Z : Y \rightrightarrows X$ between metric spaces $X$ and $Y$.

- **Local Lipschitz property** at $\bar{y}$: $\exists \, L, \varepsilon > 0$
  $$d_{Z(y_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2), \forall y_1, y_2 \in B_\varepsilon(\bar{y}).$$

- **Aubin property** at $\bar{y}$: $\exists \, L, \varepsilon > 0$
  $$d_{Z(y_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2) \cap B_\varepsilon(\bar{x}), \forall y_1, y_2 \in B_\varepsilon(\bar{y}),$$
  where $\bar{x} \in Z(\bar{y})$.

- **Local upper Lipschitz property** at $\bar{y}$: $\exists \, L, \varepsilon > 0$
  $$d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(\bar{y}), \forall y \in B_\varepsilon(\bar{y}).$$

- **Calmness** at $(\bar{y}, \bar{x}) \in \text{Gph} \ Z$: $\exists \, L, \varepsilon > 0$
  $$d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(\bar{y}) \cap B_\varepsilon(\bar{x}), \forall y \in B_\varepsilon(\bar{y}).$$
Lipschitz-like properties of set-valued mappings

Set valued mapping (multifunction) $Z : Y \Rightarrow X$ between metric spaces $X$ and $Y$.

- **Local Lipschitz property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(y_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2), \forall y_1, y_2 \in B_\varepsilon(\bar{y}).$$

- **Aubin property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(y_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2) \cap B_\varepsilon(\bar{x}), \forall y_1, y_2 \in B_\varepsilon(\bar{y}),$$
  
  where $\bar{x} \in Z(\bar{y})$.

- **Local upper Lipschitz property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(y), \forall y \in B_\varepsilon(\bar{y}).$$

- **Calmness** at $(\bar{y}, \bar{x}) \in \text{Gph } Z$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(y) \cap B_\varepsilon(\bar{x}), \forall y \in B_\varepsilon(\bar{y}).$$
Set valued mapping (multifunction) $Z : Y \rightarrow X$ between metric spaces $X$ and $Y$.

- **Local Lipschitz property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(\bar{y}_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2), \forall y_1, y_2 \in B_{\varepsilon}(\bar{y}).$$

- **Aubin property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(\bar{y}_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2) \cap B_{\varepsilon}(\bar{x}), \forall y_1, y_2 \in B_{\varepsilon}(\bar{y}),$$
  where $\bar{x} \in Z(\bar{y})$.

- **Local upper Lipschitz property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(y), \forall y \in B_{\varepsilon}(\bar{y}).$$

- **Calmness** at $(\bar{y}, \bar{x}) \in \text{Gph } Z$: $\exists L, \varepsilon > 0$
  
  $$d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(y) \cap B_{\varepsilon}(\bar{x}), \forall y \in B_{\varepsilon}(\bar{y}).$$
Lipschitz-like properties of set-valued mappings

Set valued mapping (multifunction) $Z : Y \to X$ between metric spaces $X$ and $Y$.

- **Local Lipschitz property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  \[ d_{Z(y_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2), \forall y_1, y_2 \in B_\varepsilon(\bar{y}). \]

- **Aubin property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  \[ d_{Z(y_1)}(x) \leq L \cdot d(y_1, y_2), \forall x \in Z(y_2) \cap B_\varepsilon(\bar{x}), \forall y_1, y_2 \in B_\varepsilon(\bar{y}), \]
  where $\bar{x} \in Z(\bar{y})$.

- **Local upper Lipschitz property** at $\bar{y}$: $\exists L, \varepsilon > 0$
  \[ d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(\bar{y}), \forall y \in B_\varepsilon(\bar{y}). \]

- **Calmness** at $(\bar{y}, \bar{x}) \in \text{Gph } Z$: $\exists L, \varepsilon > 0$
  \[ d_{Z(\bar{y})}(x) \leq L \cdot d(\bar{y}, y), \forall x \in Z(y) \cap B_\varepsilon(\bar{x}), \forall y \in B_\varepsilon(\bar{y}). \]
Let $F : X \rightarrow Y$ be a **continuous** mapping, $\Lambda \subseteq Y$ be a closed set. Denote by

$$M(u) = \{ x \in X : F(x) + u \in \Lambda \}$$

the perturbation of the (original) constraint set $M(0) = F^{-1}(\Lambda)$. 
Hoheisel et al. (2010), Proposition 3.5:

**Proposition**

Let $x^*$ be a local minimizer such that $M$ is calm at $(0, x^*)$. Then the original problem is calm at $x^*$. 
global: $M$ is calm at each point of its graph whenever this graph is polyhedral, cf. Rockafellar, Wets (2003), Example 9.57.

local: **Generalized Mangasarian-Fromowitz constraint qualification** (GMFCQ) for continuously differentiable $f$ and $F$:

$$F'(x^*)\lambda = 0 \& \lambda \in N_\Lambda(F(x^*)) \Rightarrow \lambda = 0.$$  

Limiting normal cone to $\Pi$ at $a$ is defined by

$$N_\Pi(a) = \limsup_{a' \to a} \hat{N}_\Pi(a), \text{ where } \Pi \subseteq \mathbb{R}^p, a \in \text{cl } \Pi. \text{ Fréchet normal cone to } \Pi \text{ at } a$$

$$\hat{N}_\Pi(a) = \left\{ \xi \in \mathbb{R}^p : \limsup_{a' \to a} \frac{\langle \xi, a' - a \rangle}{\|a - a'\|} \right\}$$

• further generalization of GMFCQ using kernels of coderivatives ...
Clarke’s exact penalty method

Clarke (1983), Proposition 6.3.1, Proposition 6.4.2, Proposition 6.4.3:

Exact $L_1$ penalization of inequality constraints can be obtained under the following assumptions (which imply “calmness” of the problem):

- inequality constraints $g_j$ are convex,
- $X$ is convex bounded,
- Slater condition (existence of a strictly feasible point),
- objective function $f$ is Lipschitz on $X$.

Employed by Meskarial et al. (2012), Sun et al. (2013).
Contents

1. Exact penalty method and calmness
2. Stochastic programming formulations
3. Relations between formulations
4. Sample approximations using Monte-Carlo techniques
5. References
### Stochastic programming formulations

1. **Stochastic programming formulation**
   - Program with a random factor

   \[ \rightarrow \rightarrow \rightarrow \]

   - Chance constrained problem (CCP)
   - Problem with penalty objective (PPO)
   - Integrated chance constrained problem (ICC)

2. **Sample approximation (S.A.)**
   - S.A. CCP
   - S.A. PPO
   - S.A. ICC

3. **Solution validation**
   - Reliability check
   - Reliability check
   - Reliability check
Optimization problem with a random factor

Program with a random factor $\xi$

$$
\min_{x \in X} \{ f(x) : g_i(x, \xi) \leq 0, \ i = 1, \ldots, k \},
$$

where $g_i, i = 0, \ldots, k$, are real functions on $\mathbb{R}^n \times \mathbb{R}^{n'}$, $X \subseteq \mathbb{R}^n$ and $\xi \in \mathbb{R}^{n'}$ is a realization of a $n'$-dimensional random vector defined on the probability space $(\Omega, \mathcal{F}, P)$.

If $P$ is known, we can use chance constraints to deal with the random constraints...
**Stochastic programming formulations**

**Chance constrained problem (CCP)**

**Chance constrained problem**

\[
\psi_\epsilon = \min_{x \in X} f(x),
\]

s.t.

\[
P(g_{11}(x, \xi) \leq 0, \ldots, g_{1k_1}(x, \xi) \leq 0) \geq 1 - \epsilon_1, \\
\vdots \\
P(g_{m1}(x, \xi) \leq 0, \ldots, g_{mk_m}(x, \xi) \leq 0) \geq 1 - \epsilon_m,
\]

with optimal solution \(x_\epsilon\), where we denoted \(\epsilon = (\epsilon_1, \ldots, \epsilon_m)\) with levels \(\epsilon_j \in (0, 1)\). The formulation covers the **joint** (\(k_1 > 1\) and \(m = 1\)) as well as the **individual** (\(k_j = 1\) and \(m > 1\)) chance constrained problems as special cases.
In general,
  - the feasible region is **not convex** even if the functions are convex,
  - it is even not easy to check **feasibility** because it leads to computations of multivariate integrals.

Hence, we will try to reformulate the chance constrained problem using penalty functions.
Penalty functions $\vartheta_j : \mathbb{R}^m \rightarrow \mathbb{R}_+$, $j = 1, \ldots, m$, are continuous nondecreasing, equal to 0 on $\mathbb{R}^m$ and positive otherwise, e.g.

$$
\vartheta^1_p(u) = \sum_{i=1}^{k} ([u_i]^+)^p, \quad p \in \mathbb{N}
$$

$$
\vartheta^2(u) = \max_{1\leq i \leq k} [u_i]^+,
$$

$$
= \min \left\{ t \geq 0 : u_i - t \leq 0, \quad i = 1, \ldots, k \right\}
$$

where $u \in \mathbb{R}^m$. Let $p_j$ denote the **penalized constraints**

$$
p_j(x, \xi) = \vartheta_j(g_{j1}(x, \xi), \ldots, g_{jk_j}(x, \xi)), \forall j.
$$
Penalty function problems

Problem with **penalties in the objective function**

\[
\varphi_N = \min_{x \in X} \left\{ f(x) + N \cdot \sum_{j=1}^{m} \mathbb{E}[p_j(x, \xi)] \right\}
\]

with an optimal solution \( x_N \).

Problem with **generalized integrated chance constraints**

\[
\varphi_{ICC}^L = \min_{x \in X} \left\{ f(x) : \text{s.t. } \mathbb{E}[p_j(x, \xi)] \leq L_j, j = 1, \ldots, m \right\}
\]

for some prescribed bounds \( L_j \geq 0, L = (L_1, \ldots, L_m)' \), with an optimal solution \( x_{ICC}^L \) (originally defined using \( \vartheta^2 \), cf. Klein Haneveld (1986)).
Penalty function problems

Problem with **penalties in the objective function**

\[ \varphi_N = \min_{x \in \mathcal{X}} \left\{ f(x) + N \cdot \sum_{j=1}^{m} \mathbb{E}[p_j(x, \xi)] \right\} \]

with an optimal solution \( x_N \).

Problem with **generalized integrated chance constraints**

\[ \varphi_{L}^{ICC} = \min_{x \in \mathcal{X}} \left\{ f(x) : \text{s.t. } \mathbb{E}[p_j(x, \xi)] \leq L_j, j = 1, \ldots, m \right\} \]

for some prescribed bounds \( L_j \geq 0, L = (L_1, \ldots, L_m)' \), with an optimal solution \( x_{L}^{ICC} \) (originally defined using \( \vartheta^2 \), cf. Klein Haneveld (1986)).
Stochastic programming formulations

History and applications of the penalty approach in SP

- Prékopa (1973): **CPP and penalization**
- Ermoliev et al (2000): Managing exposure to **catastrophic risks** (asymptotic equivalence with particular penalty)
- Branda and Dupačová (2008, 2012): **Contamination technique** for CCP (asymptotic equivalence using general penalty functions)
- Žampachová (2009): **Beam design** (reliability problem with partial differential equations - nonlinear - significant reduction of computational time)
- Branda (2009, 2012a): **Value at Risk optimization** with transaction costs and integer allocations (general penalty functions and several CC)
- M.B (2011): **Blending problem** (asymptotic equivalence with generalized integrated chance constraints)
1. Exact penalty method and calmness

2. Stochastic programming formulations

3. Relations between formulations

4. Sample approximations using Monte-Carlo techniques

5. References
Relations between formulations

1. Stochastic programming formulation
   - Chance constrained problem (CCP) ⇔ Problem with penalty objective (PPO) ⇔ Integrated chance constrained problem (ICC)

2. Sample approximation (S.A.)
   - S.A. CCP
   - S.A. PPO
   - S.A. ICC

3. Solution validation
   - Reliability check
   - Reliability check
   - Reliability check
Relations between formulations

**CCP under finite discrete distribution (FDD)**

Let the distribution of random vector $\xi$ be discrete with finite number of realizations $\xi^s$, $s = 1, \ldots, S$ with known probabilities $0 < p_s \leq 1$, $\sum_{s=1}^{S} p_s = 1$. The chance constrained problem can be then formulated as

$$
\varphi_{\varepsilon}^{CCP} = \min_{x \in X} f(x) \\
\text{s.t.} \\
\sum_{s=1}^{S} p_s I(g_1(x, \xi^s) \leq 0, \ldots, g_k(x, \xi^s) \leq 0) \geq 1 - \varepsilon,
$$

where $I$ denotes the indicator function which is equal to one if the condition is satisfied, and 0 otherwise.
Branda (2012a), Branda, Dupačová (2012): Under the following assumptions, the **asymptotic equivalence of the CCP and the PPO problems** can be shown:

- **Compactness** of the fixed set of feasible solutions.
- **Continuity** of the objective function, constraints and *the probabilistic functions*.
- Existence of **integrable majorants**.
- Existence of a **permanently feasible solution**.

THEN for any prescribed $\epsilon \in (0, 1)^m$ there always exists $N$ large enough so that minimization of the penalty objective generates optimal solutions $x_N$ which also satisfy the chance constraints with the given $\epsilon$. We can obtain asymptotic bounds on the optimal values.
Branda (2013a): Under the following assumptions, the asymptotic equivalence of the CCP and the PPO problems under finite discrete distributions can be shown:

- **Compactness** of the fixed set of feasible solutions.
- **Continuity** of the objective function and constraints:
  \[ g_i(\cdot, \xi^s), \ i = 1, \ldots, k \] are continuous for all \( s = 1, \ldots, S \).
- **Existence of a permanently feasible solution**:
  \[ g_i(x', \xi^s) \leq 0, \ i = 1, \ldots, k \] for all \( s = 1, \ldots, S \) for at least one \( x' \in X \).
We consider the following perturbed version of the problem with a random factor under FDD:

$$\begin{align*}
\min_{x \in X} & \quad f(x) \\
\text{s.t.} & \quad g_1(x, \xi^s) \leq u_{1s}, \ldots, g_k(x, \xi^s) \leq u_{ks}, \quad s = 1, \ldots, S.
\end{align*}$$

We define modified $L_1$-norm for a vector $u \in \mathbb{R}^{kS}$ as

$$\|u\| = \sum_{s=1}^{S} \sum_{i=1}^{k} p_s |u_{is}|,$$

which is necessary for showing the asymptotic equivalence.
Perturbed problem

We consider the following perturbed version of the problem with a random factor under FDD:

\[
\begin{align*}
\min_{x \in X} & \quad f(x) \\
\text{s.t.} & \quad g_1(x, \xi_s) \leq u_{1s}, \ldots, g_k(x, \xi_s) \leq u_{ks}, s = 1, \ldots, S.
\end{align*}
\]

We define modified $L_1$-norm for a vector $u \in \mathbb{R}^{kS}$ as

\[
\|u\| = \sum_{s=1}^{S} p_s \sum_{i=1}^{k} |u_{is}|,
\]

which is necessary for showing the asymptotic equivalence.
**Definition**

Let $x^*$ be feasible for the unperturbed problem, i.e. (2) with $u_{ks} = 0$. Then the problem is said to be calm at $x^*$ if there exist constant $\tilde{N}$ (modulus) and $\epsilon > 0$ (radius) such that for all $(x, u) \in \mathbb{R}^n \times \mathbb{R}^{kS}$ satisfying $x \in B_\epsilon(x^*)$ and $g_i(x, \xi^s) \leq u_{is}$, one has

$$f(x) + \tilde{N} \|u\| \geq f(x^*).$$

Note that then $x^*$ is necessarily a local solution to the unperturbed problem.
Proposition

Let $x^*$ be feasible for the unperturbed problem, i.e. (2) with $u_{ks} = 0$, $i = 1, \ldots, k$, $s = 1, \ldots, S$. Then the unperturbed problem is calm at $x^*$ with modulus $\tilde{N} \geq 0$ and radius $\epsilon > 0$ if and only if $x^*$ is a local minimum of the function

$$f(x) + N \sum_{s=1}^{S} p_s \sum_{i=1}^{k} |g_i(x, \xi^s)|_+$$

over $B_\epsilon(x^*)$ for all $N \geq \tilde{N}$.

The theorems on asymptotic equivalence can be modified for local minimizers, i.e. $X$ is replaced by $X \cap B_\epsilon(x^*)$ for some local optimal solution $x^*$ in the following theorems. Moreover, a special form of the penalty function is necessary: $\Phi(x, \xi^s) = \sum_{i=1}^{k} |g_i(x, \xi^s)|_+$. 
Proposition

Let \( x^* \) be feasible for the unperturbed problem, i.e. (2) with \( u_{ks} = 0 \), \( i = 1, \ldots, k \), \( s = 1, \ldots, S \). Then the unperturbed problem is calm at \( x^* \) with modulus \( \tilde{N} \geq 0 \) and radius \( \epsilon > 0 \) if and only if \( x^* \) is a local minimum of the function

\[
f(x) + N \sum_{s=1}^{S} p_s \sum_{i=1}^{k} |g_i(x, \xi^s)|_+\]

over \( B_\epsilon(x^*) \) for all \( N \geq \tilde{N} \).

The theorems on asymptotic equivalence can be modified for local minimizers, i.e. \( X \) is replaced by \( X \cap B_\epsilon(x^*) \) for some local optimal solution \( x^* \) in the following theorems. Moreover, a special form of the penalty function is necessary: \( \Phi(x, \xi^s) = \sum_{i=1}^{k} |g_i(x, \xi^s)|_+ \).
We consider the CCP and PFP problems and assume:

(i) \( g_i(x', \xi^s) \leq 0, \ i = 1, \ldots, k \) for all \( s = 1, \ldots, S \) for at least one \( x' \in X \).

(ii) the corresponding unperturbed problem, i.e. (2) with \( u_{ks} = 0 \), is calm at its local optimal solutions modulus \( \tilde{N} \geq 0 \) and radius \( \epsilon > 0 \).

For arbitrary \( \gamma \in (0, 1) \), \( N > 0 \) and \( \varepsilon \in (0, 1) \) put

\[
\varepsilon(x) = \sum_{s=1}^{S} p_s I(\Phi(x, \xi^s) > 0),
\]

\[
\alpha_N(x) = N \sum_{s=1}^{S} p_s \Phi(x, \xi^s),
\]

\[
\beta_\varepsilon(x) = \frac{1}{\varepsilon^\gamma} \sum_{s=1}^{S} p_s \Phi(x, \xi^s).
\]

Then for any prescribed \( \varepsilon \in (0, 1) \) there always exists \( N \leq \tilde{N} \) large enough so that minimization of the penalty objective generates optimal solutions \( x_N \) which also satisfy the probabilistic constraint with the given
We consider the CCP and PFP problems and assume:

(i) \( g_i(x', \xi^s) \leq 0, \ i = 1, \ldots, k \) for all \( s = 1, \ldots, S \) for at least one \( x' \in X \).

(ii) the corresponding unperturbed problem, i.e. (2) with \( u_{ks} = 0 \), is calm at its local optimal solutions modulus \( \tilde{N} \geq 0 \) and radius \( \epsilon > 0 \).

For arbitrary \( \gamma \in (0, 1) \), \( N > 0 \) and \( \epsilon \in (0, 1) \) put

\[
\epsilon(x) = \sum_{s=1}^{S} p_s I(\Phi(x, \xi^s) > 0),
\]

\[
\alpha_N(x) = N \sum_{s=1}^{S} p_s \Phi(x, \xi^s),
\]

\[
\beta_\epsilon(x) = \frac{1}{\epsilon^\gamma} \sum_{s=1}^{S} p_s \Phi(x, \xi^s).
\]

Then for any prescribed \( \epsilon \in (0, 1) \) there always exists \( N \leq \tilde{N} \) large enough so that minimization of the penalty objective generates optimal solutions \( x_N \) which also satisfy the probabilistic constraint with the given \( \epsilon \).
Moreover, bounds on the local optimal value $\psi_\varepsilon$ of based on the local optimal value $\varphi_N$ and vice versa can be constructed:

$$
\varphi_{1/\varepsilon}^{1/\gamma}(x_N) - \beta_\varepsilon(x_N)(x_{\varepsilon}^{CCP}(x_N)) \leq \psi_\varepsilon(N) \leq \varphi_N - \alpha_\varepsilon(x_N),
$$

$$
\psi_\varepsilon(x_N) + \alpha_\varepsilon(x_N) \leq \varphi_N \leq \psi_{N-1/\gamma} + \beta_{N-1/\gamma}(x_{\varepsilon}^{CCP}(x_{N-1/\gamma})),
$$

with

$$
\lim_{N \to \tilde{N}_-} \alpha_\varepsilon(x_N) = \lim_{N \to \tilde{N}_-} \varepsilon(x_N) = \lim_{\varepsilon \to \tilde{\varepsilon}_+} \lambda_\varepsilon(x_{\varepsilon}^{CCP}) = 0,
$$

for any sequences of the optimal solutions $x_N$ and $x_{\varepsilon}^{CCP}$, where $\tilde{\varepsilon} < \min_s \rho_s$. 
Remark

If we want to obtain exact convergence of the bounds on optimal values, we have to add the following condition. Since $N^{-1/\gamma}$ in the upper bound on $\varphi_N$ has to converge to $\tilde{\epsilon}$, we obtain the condition $N \geq \max\{\tilde{N}, \tilde{\epsilon}^{-\gamma}\}$. 
Contents

1 Exact penalty method and calmness
2 Stochastic programming formulations
3 Relations between formulations
4 Sample approximations using Monte-Carlo techniques
5 References
Sample approximations using Monte-Carlo techniques

Program with a random factor

Stochastic programming formulation


Sample approximation (S.A.)

2. S.A. CCP S.A. PPO S.A. ICC

Solution validation

3. Reliability check Reliability check Reliability check
Sample approximations using Monte-Carlo techniques

The SAA technique for expected value constrained problems

- Wang, Ahmed (2008): one constraint, iid sampling, finite or infinite bounded set of feasible solutions and Lipschitz continuity
- Branda (2012b): several constraints, iid sampling, finite or infinite bounded set of feasible solutions and Lipschitz continuity (fixed or random modulus)
- Branda (2013c): several constraints, non-iid sampling, mixed-integer set of feasible solutions and H-calmness

Examples: (generalized) integrated chance constraints, risk-shaping with CVaRs, diversification-consistent DEA tests with CVaR deviations and limited diversification etc., cf. Branda (2012c, 2013b, 2013c).
Sample approximations using Monte-Carlo techniques

The SAA technique for expected value constrained problems

- Wang, Ahmed (2008): one constraint, iid sampling, finite or infinite bounded set of feasible solutions and Lipschitz continuity
- Branda (2012b): several constraints, iid sampling, finite or infinite bounded set of feasible solutions and Lipschitz continuity (fixed or random modulus)
- Branda (2013c): several constraints, non-iid sampling, mixed-integer set of feasible solutions and H-calmness

Examples: (generalized) integrated chance constraints, risk-shaping with CVaRs, diversification-consistent DEA tests with CVaR deviations and limited diversification etc., cf. Branda (2012c, 2013b, 2013c).
The SAA technique for expected value constrained problems

- Wang, Ahmed (2008): one constraint, iid sampling, finite or infinite bounded set of feasible solutions and Lipschitz continuity
- Branda (2012b): several constraints, iid sampling, finite or infinite bounded set of feasible solutions and Lipschitz continuity (fixed or random modulus)
- Branda (2013c): several constraints, non-iid sampling, mixed-integer set of feasible solutions and H-calmness

Examples: (generalized) integrated chance constraints, risk-shaping with CVaRs, diversification-consistent DEA tests with CVaR deviations and limited diversification etc., cf. Branda (2012c, 2013b, 2013c).
Let $X \subseteq \mathbb{R}^n$, $Y \subseteq \mathbb{Z}^{n'}$ and

$$Z = \{(x, y) \in X \times Y : h_1(x, y) \leq 0, \ldots, h_k(x, y) \leq 0\}$$

be the deterministic mixed-integer part of the set of feasible solutions with $h_j(x, y) : \mathbb{R}^n \times \mathbb{R}^{n'} \rightarrow \mathbb{R}$. Let $\xi$ be a random vector on the probability space $(\Omega, \mathcal{F}, P)$, $p_j(x, y, \xi), j = 1, \ldots, m$, be real functions on $\mathbb{R}^n \times \mathbb{R}^{n'} \times \mathbb{R}^{n''}$ measurable in $\xi$ for all $x \in X$ and $y \in Y$. We assume that the objective function $f(x, y) : \mathbb{R}^n \times \mathbb{R}^{n'} \rightarrow \mathbb{R}$ does not depend on the random vector.
We denote the set of feasible solutions as

\[ Z_L = \left\{ (x, y) \in Z : \, p_j(x, y) := \mathbb{E}[p_j(x, y, \xi)] \leq L_j, \, j = 1, \ldots, m \right\} \]

for some prescribed bounds \( L_j \in \mathbb{R}, \, L = (L_1, \ldots, L_m)' \). We assume that the levels are chosen in such a way that the set of feasible solutions is nonempty and that the expectations are finite for all \((x, y) \in Z\). Then, the stochastic programming problem with the expected value constraints can be formulated as

\[ \min_{(x, y) \in Z_L} f(x, y). \]
Let $\xi^1, \ldots, \xi^S$ be a Monte-Carlo sample of the underlying distribution of the random vector $\xi$. We denote the set of feasible solutions of the sample-approximated problem as

$$Z^S_L = \left\{ (x, y) \in Z : p^S_j(x, y) := \frac{1}{S} \sum_{s=1}^{S} p_j(x, y, \xi^s) \leq L_j, \ j = 1, \ldots, m \right\}.$$  

The sample version of the problem with several expected value constraints is defined as

$$\min_{(x, y) \in Z^S_L} f(x, y). \tag{5}$$

where the levels $L_j \in \mathbb{R}$ are allowed to be different from the original levels.
Calm functions

- \( p(x, \xi) \) is said to be be **Hölder-calm** (H-calm) at \( x \) with modulus \( M_x(\xi) > 0 \), and order \( \gamma_x > 0 \), if there exist a measurable function \( M_x(\xi) : \Xi \to \mathbb{R}_+ \) and positive numbers \( \gamma_x, \delta_x \) such that
  \[
  |p(x, \xi) - p(x', \xi)| \leq M_x(\xi) \|x - x'\|^{\gamma_x},
  \]
  for all \( \xi \in \Xi \) and \( \|x - x'\| \leq \delta_x \).

- \( p(x, \xi) \) is said to be be **almost Hölder-calm** (almost H-calm) at \( x \) with modulus \( M_x(\xi) > 0 \), and order \( \gamma_x > 0 \), if for any \( \varepsilon > 0 \), there exist a measurable function \( M_x(\xi) : \Xi \to \mathbb{R}_+ \) and positive numbers \( \gamma_x, \delta_x, C \) and an open set \( \Delta_x(\varepsilon) \subset \Xi \) such that
  \[
P(\xi \in \Delta_x(\varepsilon)) \leq C \varepsilon
  \]
  and
  \[
  |p(x, \xi) - p(x', \xi)| \leq M_x(\xi) \|x - x'\|^{\gamma_x},
  \]
  for all \( \xi \in \Xi \setminus \Delta_x(\varepsilon) \) and \( \|x - x'\| \leq \delta_x \).
Calm functions

- \( p(x, \xi) \) is said to be be Hölder-calm (H-calm) at \( x \) with modulus \( M_x(\xi) > 0 \), and order \( \gamma_x > 0 \), if there exist a measurable function \( M_x(\xi) : \Xi \rightarrow \mathbb{R}_+ \) and positive numbers \( \gamma_x, \delta_x \) such that

\[
|p(x, \xi) - p(x', \xi)| \leq M_x(\xi) \|x - x'\|^\gamma_x,
\]

for all \( \xi \in \Xi \) and \( \|x - x'\| \leq \delta_x \).

- \( p(x, \xi) \) is said to be be almost Hölder-calm (almost H-calm) at \( x \) with modulus \( M_x(\xi) > 0 \), and order \( \gamma_x > 0 \), if for any \( \varepsilon > 0 \), there exist a measurable function \( M_x(\xi) : \Xi \rightarrow \mathbb{R}_+ \) and positive numbers \( \gamma_x, \delta_x, C \) and an open set \( \Delta_x(\varepsilon) \subset \Xi \) such that

\[
P(\xi \in \Delta_x(\varepsilon)) \leq C\varepsilon
\]

and

\[
|p(x, \xi) - p(x', \xi)| \leq M_x(\xi) \|x - x'\|^\gamma_x,
\]

for all \( \xi \in \Xi \setminus \Delta_x(\varepsilon) \) and \( \|x - x'\| \leq \delta_x \).
First we review a result which holds if \( n = 0 \) and \( m = 1 \), i.e. the set of feasible solutions is finite with one expected value constraint. Then, see the proof of Proposition 1 by Wang and Ahmed (2008) under iid sampling, it holds that

\[
P(Z^S_{L+\tau} \subseteq Z_L) \geq 1 - |Y| \exp \left\{ - S \min_{y \in Y} I_y(\tau) \right\},
\]

where \(|Y|\) denotes the cardinality of \( Y \), \( I_y \) is the large deviation rate functions, i.e. the Fenchel dual to the logarithm of the finite moment generating function of the difference \( p(y, \xi) - p(y) \) which is defined as

\[
I_y(\tau) = \sup_{t \in \mathbb{R}} \left\{ t \tau - \Psi_y(t) \right\},
\]

where

\[
\Psi_y(t) = \ln \mathbb{E} \left[ e^{t(p(y, \xi) - p(y))} \right], \quad p(y) = \mathbb{E}[p(y, \xi)].
\]

The estimate is based on Cramér’s large deviation theory, see, e.g., Dembo and Zeitouni (2010).
The estimate remains valid also for **non-iid sampling** if the **Gärtner-Ellis theorem** is used and a condition on the convergence of logarithmic moment generating functions is added. For every $y \in Y$ and $t \in \mathbb{R}$, denote

$$
\Psi^S_y(t) = \ln \mathbb{E}\left[e^{t(p^S(y) - p(y))}\right].
$$

Then the assumption of Gärtner-Ellis theorem holds if

$$
\Psi_y(t) = \lim_{S \to \infty} \frac{\Psi^S_y(St)}{S} \quad (7)
$$

exists as an extended real number for all $t \in \mathbb{R}$ and $\Psi_y(t) < \infty$ for $t$ close to 0, c.f. Theorem 2.3.6 in Dembo and Zeitouni (2010). This conditions is trivially fulfilled for iid samples. Moreover, it can be verified for **finite state Markov chains**, see Chapter 3 in Dembo and Zeitouni (2010). See also Drew, Homem-de-Mello (2012) for **Latin Hypercube Sampling**.
Branda (2013c): Let

(i) $Y \subseteq \mathbb{R}^{n'}$ be finite, and $X \subseteq \mathbb{R}^n$ be bounded, i.e. $D = \sup \{ \| x - x' \|_{\infty} : x, x' \in X \}$ be a finite diameter,

(ii) $p(x, y, \xi)$ be uniformly H-calm in $x \in X$ for each $y \in Y$, moduli $M(\xi) > 0$, and order $\gamma > 0$:

$$|p(x, y, \xi) - p(x', y, \xi)| \leq M(\xi) \| x - x' \|^{\gamma}, \forall x, x' \in X, \forall \xi \in \Xi,$$

with $M = \mathbb{E}[M(\xi)] < \infty$,

(iii) the logarithmic moment generating functions $\Psi_{xy}(t)$ of $p(x, y, \xi) - p(x, y)$ be finite around 0 and

$$\Psi_{xy}(t) = \lim_{S \to \infty} \frac{\Psi_{xy}^{S}(St)}{S},$$

for all $t \in \mathbb{R}$ and for all $(x, y) \in Z$.

(iv) the logarithmic moment generating function $\Psi_{M}(t)$ of $M(\xi) - M$ be finite around 0 and

$$\Psi_{M}(t) = \lim_{S \to \infty} \frac{\Psi_{M}^{S}(St)}{S}, \forall t \in \mathbb{R}.$$
Branda (2013c): Let

(i) \( Y \subseteq \mathbb{R}^{n'} \) be finite, and \( X \subseteq \mathbb{R}^n \) be bounded, i.e. \( D = \sup \{ \| x - x' \|_{\infty} : x, x' \in X \} \) be a finite diameter,

(ii) \( p(x, y, \xi) \) be uniformly H-calm in \( x \in X \) for each \( y \in Y \), moduli \( M(\xi) > 0 \), and order \( \gamma > 0 \):

\[
|p(x, y, \xi) - p(x', y, \xi)| \leq M(\xi) \| x - x' \|^{\gamma} \quad \forall x, x' \in X, \forall \xi \in \Xi,
\]

with \( M = \mathbb{E}[M(\xi)] < \infty \),

(iii) the logarithmic moment generating functions \( \Psi_{xy}(t) \) of \( p(x, y, \xi) - p(x, y) \) be finite around 0 and

\[
\Psi_{xy}(t) = \lim_{S \to \infty} \frac{\Psi^S_{xy}(St)}{S},
\]

for all \( t \in \mathbb{R} \) and for all \( (x, y) \in Z \).

(iv) the logarithmic moment generating function \( \Psi_M(t) \) of \( M(\xi) - M \) be finite around 0 and

\[
\Psi_M(t) = \lim_{S \to \infty} \frac{\Psi^S_M(St)}{S}, \forall t \in \mathbb{R}.
\]
Sample approximations using Monte-Carlo techniques

Branda (2013c): Let

(i) $Y \subseteq \mathbb{R}^{n'}$ be finite, and $X \subseteq \mathbb{R}^n$ be bounded, i.e. $D = \sup \{\|x - x'\|_\infty : x, x' \in X\}$ be a finite diameter,

(ii) $p(x, y, \xi)$ be uniformly H-calm in $x \in X$ for each $y \in Y$, moduli $M(\xi) > 0$, and order $\gamma > 0$:

$$|p(x, y, \xi) - p(x', y, \xi)| \leq M(\xi) \|x - x'\|_\gamma, \ \forall x, x' \in X, \ \forall \xi \in \Xi,$$

with $M = \mathbb{E}[M(\xi)] < \infty$,

(iii) the logarithmic moment generating functions $\Psi_{xy}(t)$ of $p(x, y, \xi) - p(x, y)$ be finite around 0 and

$$\Psi_{xy}(t) = \lim_{S \to \infty} \frac{\Psi^S_{xy}(St)}{S},$$

for all $t \in \mathbb{R}$ and for all $(x, y) \in Z$.

(iv) the logarithmic moment generating function $\Psi_M(t)$ of $M(\xi) - M$ be finite around 0 and

$$\Psi_M(t) = \lim_{S \to \infty} \frac{\Psi^S_M(St)}{S}, \ \forall t \in \mathbb{R}.$$
Then, for $\tau > 0$ small,

(a) the probability that the set of feasible solutions is contained in the relaxed sample-approximated set of feasible solutions increases exponentially with increasing sample size, and it holds

$$P(Z_L \subseteq Z_{L+\tau}^S) \geq 1 - \left(1 + |Y| \frac{D^n(4M + \tau)^{n/\gamma}}{\tau^{n/\gamma}}\right) \exp \left\{ - Sd(\tau) \right\},$$

where

$$\sigma_{xy}^2 = \text{Var}[p(x, y, \xi) - p(x, y)],$$

$$\sigma_M^2 = \text{Var}[M(\xi) - M],$$

$$\nu = \left(\frac{\tau}{4M + \tau}\right)^{1/\gamma},$$

$$d(\tau) = \min \left\{ \min_{(x,y) \in Z_\nu} \frac{\tau^2}{8\sigma_{xy}^2}, \frac{\tau^2}{8\sigma_M^2} \right\}.$$
(b) we can get an estimate for the sample size which is necessary to ensure that the original feasibility set is contained in the relaxed sample-approximated feasibility set with a high probability, equal to $1 - \delta$, $\delta \in (0, 1)$:

\[
S \geq \frac{1}{d(\tau)} \left( \ln \frac{1}{\delta} + \ln \left( 1 + |Y| \frac{D^n(4M + \tau)^{n/\gamma}}{\tau^{n/\gamma}} \right) \right).
\]
We can compare our estimate for $\gamma = 1$ with the result of Wang and Ahmed (2008) that is also valid for our problem with iid sampling, but it does not take into account the structure of the set of feasible solutions. In our case, H-calmness (Lipschitz continuity) is necessary only with respect to the continuous variables and the diameter of the set relates to the real bounded part only. To apply Wang and Ahmed (2008) estimate, Lipschitz modulus $\tilde{M}(\xi) > 0$ of function $p(\cdot, \cdot, \xi)$ is necessary, i.e. for $\xi \in \Xi$

$$|p(x, y, \xi) - p(x', y', \xi)| \leq \tilde{M}(\xi) \|(x, y) - (x', y')\|_\infty, \, \forall (x, y), (x', y') \in \mathcal{Z}.$$  

The Lipschitz constants are then incorporated into the rate estimate $d(\tau)$. Higher the variance of the random Lipschitz modulus is, lower the rate of convergence and higher the sample size estimate are obtained.
Moreover, the difference can be also identified in the estimate of the size of the finite set, which is necessary to approximate the set of feasible solutions. In our case, it is

\[ |Y| \left( \frac{D(4M + \tau)}{\tau} \right)^n. \]

In the proof of Proposition 2 by Wang and Ahmed (2008), it was estimated by

\[ \left( \frac{\tilde{D}(4\tilde{M} + \tau)}{\tau} \right)^{n+n'}, \]

where the diameter \( \tilde{D} \) is computed with respect to the continuous and discrete part of the set of feasible solutions

\[ \tilde{D} = \sup \{ \|(x, y) - (x', y')\|_\infty : (x, y), (x', y') \in Z \} \]

and \( \tilde{M} = \mathbb{E}\tilde{M}(\xi). \)
Branda (2013c): Extension to several expected value constraints, comparison of the estimates on a problem with CVaR risk-shaping.
1. Exact penalty method and calmness
2. Stochastic programming formulations
3. Relations between formulations
4. Sample approximations using Monte-Carlo techniques
5. References


