# Expected Residual Minimization for Stochastic Variational Inequalities

Xiaojun Chen

Hong Kong Polytechnic University

SJOM 2011, 26-29 September 2011

Joint work with H. Fang, M. Fukushima, G. Lin, A. Sumalee,

R.J-B Wets, C. Zhang, Y. Zhang

# Outline

- Stochastic complementarity problems
- Stochastic variational inequalities
- Smoothing sample average approximation (SSAA)
- Traffic equilibrium assignment

# I. Stochastic complementarity problems

Nonlinear complementarity problem (NCP): Given $F : R^n \to R^n$,

$$x \geq 0, \quad F(x) \geq 0, \quad x^T F(x) = 0.$$

The NCP can be reformulated as a system of nonlinear equations

$$\Phi(x, F(x)) = \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = 0$$

or a minimization problem

$$\min_{x \in R^n} \|\Phi(x, F(x))\|^2$$

by using an NCP function $\phi$, e.g.

$$\phi(x_i, F_i(x)) = \min(x_i, F_i(x)).$$

# NCP functions

A function $\phi : R^2 \to R$ is called an NCP-function if

$$\phi(a,b) = 0 \quad \Leftrightarrow \quad ab = 0, a \geq 0, b \geq 0.$$

Example of NCP functions

| | |
|---|---|
| $\phi_{NR}(a,b) = \min(a,b)$ | natural residual |
| $\phi_{FB}(a,b) = a + b - \sqrt{a^2 + b^2}$ | Fischer-Burmeister function |
| $\phi_{CCK}(a,b) = \lambda\phi_{FB}(a,b) + (1-\lambda)a_+b_+$ | penalized FB function |

Smoothing Newton methods and semismooth Newton methods are efficient to solve the NCP via the nonsmooth equations $\Phi(x, F(x)) = 0$ or minimization problem $\min \|\Phi(x, F(x))\|^2$.

Cottle-Pang-Stone (1992), Facchinei-Pang (2000), Ferris-Pang (1997), Ralph (1994), B.Chen-Harker (1997), C.Chen-Mangasarian (1996), Chen-Qi-Sun (1998), Chen-Ye (1999), B.Chen-Chen-Kanzwo (2000), Luo-Tseng(1997), Yamashita-Fukushima (1997), Qi-Sun (1993), Fukushima (2001), Han-Xiu-Qi (2006), Hu-Huang-J.Chen (2009), et al.

# Deterministic formulation using NCP function

Stochastic NCP: Given $F : \Xi \times R^n \to R^n$, find $x \in R^n$ such that

$$x \geq 0, \quad F(\xi, x) \geq 0, \quad x^T F(\xi, x) = 0, \quad \text{for} \quad \xi \in \Xi.$$

- Expected value (EV) formulation
  Gürkan-Özge-Robinson(1999), Ruszczynski-Shapiro(2003), Jiang-Xu(2008)

$$x \geq 0, \quad E[F(\xi, x)] \geq 0, \quad x^T E[F(\xi, x)] = 0$$

$$\Leftrightarrow \qquad \min_{x \in R^n} \|\Phi(x, E[F(\xi, x)])\|^2$$

- Expected residual minimization (ERM) formulation
  Chen-Fukushima(2005)

$$\min_{x \geq 0} E[\|\Phi(x, F(\xi, x))\|^2]$$

# ERM formulation for NCP

- Expected residual minimization (ERM) formulation

$$\min_{x\geq 0} \varphi(x) := E[\|\min(x, F(\xi, x))\|^2] \qquad \text{(ERM)}$$

  Chen-Fukushima(MOR 2005)

- Smoothing algorithms for solving ERM
  Chen-Zhang-Fukushima(MP 2009, one of the top 8 most cited articles published in MP in 2009-2010)
  Zhang-Chen(SIOPT 2009).
- Applications in traffic assignment
  Zhang-Chen-Sumalee (TRB 2011)
- Math. Programming with stochastic equilibrium constraints
  Lin-Chen-Fukushima (MP 2009)
- Error bounds

$$E[\text{dist}(x - X^*_\xi)] \leq \alpha E[\|\min(x, F(\xi, x))\|^2]$$

  Chen-Xiang (MP 2006, 2011, SIOPT 2007).

# II. Stochastic variational inequalities

Variational inequalities (VI): Given a closed and convex set $X$ and a continuous function $F : R^n \to R^n$, find $x \in X$ such that

$$(y - x)^T F(x) \geq 0, \quad \forall y \in X.$$

The VI can be reformulated as a minimization problem by using a residual function $f$:

(i) $f(x) \geq 0, \quad \forall x \in D \supseteq X.$
(ii) $f(x^*) = 0 \quad \Leftrightarrow \quad x^*$ solves the VI.

Projection function

$$\min_{x \in R^n} f(x) := \|x - \mathrm{Proj}_X(x - F(x))\|^2$$

Gap function

$$\min_{x \in X} f(x) := \max_{y \in X} (x - y)^T F(x).$$

# A residual function of stochastic VI

**Stochastic VI**

Given $F : \Xi \times R^n \to R^n$, $X_\xi \subset R^n$ and $\Xi \subseteq R^L$, a set representing future states of knowledge, find $x \in X_\xi$ such that

$$(y - x)^T F(\xi, x) \geq 0, \quad \forall y \in X_\xi, \qquad \xi \in \Xi.$$

**Definition of a residual function** Chen-Wets-Zhang(2011)

Let $D \subseteq R^n$ be a closed and convex set. $f : \Xi \times D \to R_+$ is a residual function of the stochastic VI, if the following conditions hold,

(i) For any $x \in D$, $\mathrm{prob}\{ f(\xi, x) \geq 0\} = 1$.

(ii) $\exists\, u : \Xi \times D \to R^n$ such that for any $x \in D$ and almost every $\xi \in \Xi$, $f(\xi, x) = 0$ if and only if $u(\xi, x)$ solves the VI$(X_\xi, F(\xi, \cdot))$.

**Example: Projection function**

$$f(\xi, x) := \|x - \mathrm{Proj}_{X_\xi}(x - F(\xi, x))\|^2$$

with $D = R^n$ and $u(\xi, x) = x$.

# Stochastic VI with linear constraints

Given $F : \Xi \times R^n \to R^n$, $X_\xi \subset R^n$ and $\Xi \subseteq R^L$, find $x \in X_\xi$ such that

$$(y-x)^T F(\xi,x) \geq 0, \quad \forall y \in X_\xi, \qquad \xi \in \Xi.$$

$$X_\xi = \{x |\, Ax = b_\xi,\, x \geq 0\}$$

Gap function for a fixed $\xi$

$$\begin{aligned} f(\xi,x) &:= \max_{y\in X_\xi} (x-y)^T F(\xi,x) \\ &= x^T F(\xi,x) + \max\{-y^T F(\xi,x) \mid Ay = b_\xi,\, y \geq 0\} \\ &= x^T F(\xi,x) + \min\{z^T b_\xi \mid A^T z + F(\xi,x) \geq 0\}. \end{aligned}$$

A residual function of stochastic VI Chen-Wets-Zhang(2011)

$$f(\xi,x) = u(\xi,x)^T F(\xi,u(\xi,x)) + \min\{z^T b_\xi \mid A^T z + F(\xi,u(\xi,x)) \geq 0\}$$

where $u(\xi,x) = x + A^\dagger(b_\xi - Ax), \qquad A^\dagger = A^T(AA^T)^{-1}$.

# Stochastic VI using a residual function

$$f(\xi,x) = u(\xi,x)^T F(\xi,u(\xi,x)) + Q(\xi,u(\xi,x))$$

$u(\xi,x) = x + A^\dagger(b_\xi - Ax)$ is 'recourse' solution (projection of $x$ on $X_\xi$).

$$\begin{aligned} Q(\xi,u(\xi,x)) &= \min\{\, z^T b_\xi \mid A^T z + F(\xi,u(\xi,x)) \geq 0\} \\ &= \max\{-y^T F(\xi,u(\xi,x)) \mid \ y \in X_\xi\}. \end{aligned}$$

Let $D = \{x|(A^\dagger A - I)x \leq c\}, \quad c_i \leq \min_{\xi\in\Xi}(A^\dagger b_\xi)_i$. Then we have $Au(\xi,x) = b_\xi$, $u(\xi,x) \geq 0$ for $x \in D \quad \Rightarrow \quad u(\xi,x) \in X_\xi$, for $x \in D$.

$$\begin{aligned} f(\xi,x) &= u(\xi,x)^T F(\xi,u(\xi,x)) + Q(\xi,u(\xi,x)) \\ &= u(\xi,x)^T F(\xi,u(\xi,x)) - y(\xi,x)^T F(\xi,u(\xi,x)) \\ &= \max\{(u(\xi,x) - y)^T F(\xi,u(\xi,x)) \,|\, y \in X_\xi\} \\ &\geq 0. \end{aligned}$$

Hence, we obtain $\mathrm{prob}\{f(\xi,x) \geq 0\} = 1$. Moreover, $f(\xi,x) = 0$ if and only if $u(\xi,x)$ solves the VI($X_\xi, F(\xi,\cdot)$) a.s

# ERM formulation vs EV formulation

- Expected residual minimization (ERM) formulation

$$\min_{x \in D} E[f(\xi, x)] = E[u(\xi, x)^T F(\xi, u(\xi, x)) + Q(\xi, u(\xi, x))]$$

$x$ is the first level decision, $u(\xi, x)$ is the recourse variable. $u(\xi, x)$ is feasible but not necessarily optimal, i.e.

$$u(\xi, x) \in X_\xi \quad \text{but} \quad f(\xi, x) \geq 0.$$

The cost function $f(\xi, x)$ measures the loss at the event $\xi$ and decision $x$. The ERM formulation minimizes the expected values of the loss for all possible occurrences due to failure of the equilibrium.

- Expected value (EV) formulation
Find $x \in \bar{X} = \{ x \mid Ax = E[b_\xi], \quad x \geq 0\}$ such that

$$(y - x)^T E[F(\xi, x)] \geq 0, \quad \forall y \in \bar{X}$$

# III: Smoothing sample average approximation

- Definition 1: Let $\varphi : R^n \to R$ be locally Lipschitz. We call $\tilde{\varphi} : R^n \times R_+ \to R$ a smoothing function of $\varphi$, if $\tilde{\varphi}(\cdot, \mu)$ is continuously differentiable in $R^n$ for any fixed $\mu > 0$, and for any $x \in R^n$,

$$\lim_{z \to x, \mu \downarrow 0} \tilde{\varphi}(z, \mu) = \varphi(x).$$

- Subdifferential associated with $\tilde{\varphi}$

$$G_{\tilde{\varphi}}(x) = \{v \ : \quad \nabla_x \tilde{\varphi}(x^\nu, \mu_\nu) \ \to v, \text{ for } x^\nu \to x, \ \mu_\nu \downarrow 0 \}.$$

Rockafellar and Wets (1998): $G_{\tilde{\varphi}}(x)$ is nonempty and bounded,

$$\partial \varphi(x) = \text{co}\{ \lim_{\substack{x_i \to x \\ x_i \in D_\varphi}} \nabla \varphi(x_i)\} \subseteq \text{co} G_{\tilde{\varphi}}(x).$$

In our problems: $\partial \varphi(x) = \text{co} G_{\tilde{\varphi}}$

# Smoothing sample average approximation(SSAA)

Xiaojun Chen, Roger J-B Wets and Yanfang Zhang (2011)

$$\text{ERM} \qquad \min_{x\in D} \varphi(x) = E[f(\xi, x)] \tag{1}$$

$$\text{smoothing ERM} \qquad \min_{x\in D} \varphi_\mu(x) = E[f(\xi, x, \mu)] \tag{2}$$

$$\text{SSAA} - \text{ERM} \qquad \min_{x\in D} \Phi_\mu^N(x) := \frac{1}{N}\sum_{i=1}^{N} \tilde{f}(\xi^i, x, \mu), \tag{3}$$

where $\tilde{f} : \Xi \times R^n \times R_+ \to R_+$ is a smoothing approximation of $f$.

$\bar{x}$ is called a stationary point of (3) if

$$\Phi_\mu^N(\bar{x}; z - \bar{x}) \geq 0, \quad \forall z \in D.$$

# Assumptions and Properties

**A1.** $X_\xi = \{x | Ax = b_\xi, x \geq 0\}$ is bounded (applications O.K.)

**A2.** $b_\xi$ and $F(\xi, x)$ are bounded for $x \in X_\xi$ and $\xi \in \Xi$, a.s. (standard)

**P1** Relatively complete recourse
recourse variable $u(\xi, x)$ is bounded and

$$\max\{-y^T F(\xi, u(\xi, x)) \mid y \in X_\xi\}$$

has a solution a.s.

**P2** $f(\xi, \cdot)$ is global Lipschitz a.s.

**P3** $E[f(\xi, \cdot)]$ is globally Lipschitz on $D \supseteq X_\xi$, and semismooth.

# Convergence of SSAA

Xiaojun Chen, Roger J-B Wets and Yanfang Zhang (2011)

Let $S_\mu^N$ and $T_\mu^N$ be the sets of solutions and stationary points of (3), respectively.

Under assumptions (A1)-(A2), if the sample is iid, then the following hold.

(1.1) Any sequence $\{x_\mu^N \in S_\mu^N\}$ has a cluster point as $N \to \infty$ and $\mu \downarrow 0$ a.s.

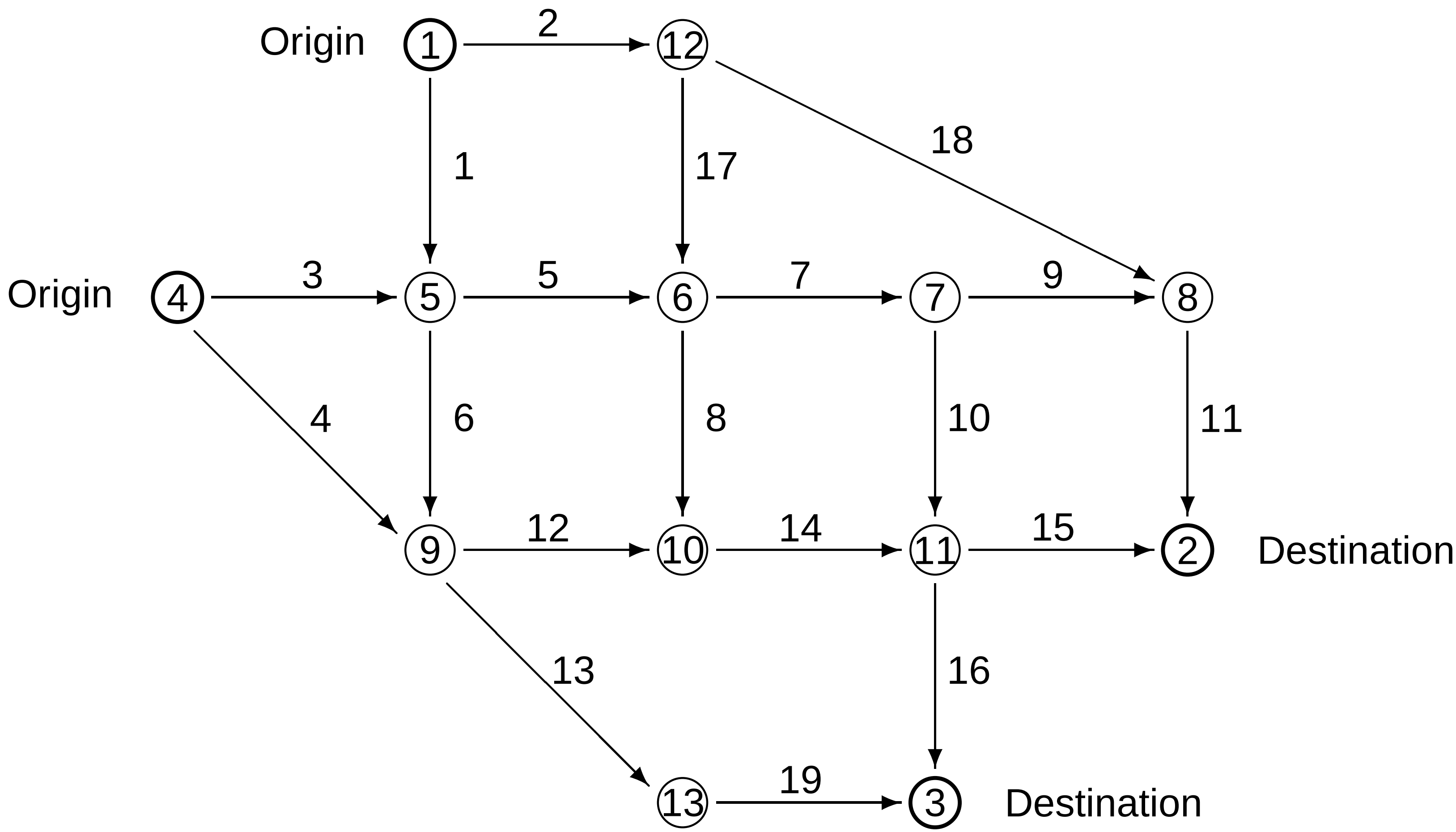
(1.2) Any cluster point of $\{x_\mu^N \in S_\mu^N\}$ is an optimal solution of the ERM (1) a.s.

(2.1) Any sequence $\{x_\mu^N \in T_\mu^N\}$ has a cluster point as $N \to \infty$ and $\mu \downarrow 0$ a.s.

(2.2) Any cluster point of $\{x_\mu^N \in T_\mu^N\}$ is a stationary point of the ERM (1) a.s.

# IV Traffic equilibrium assignment

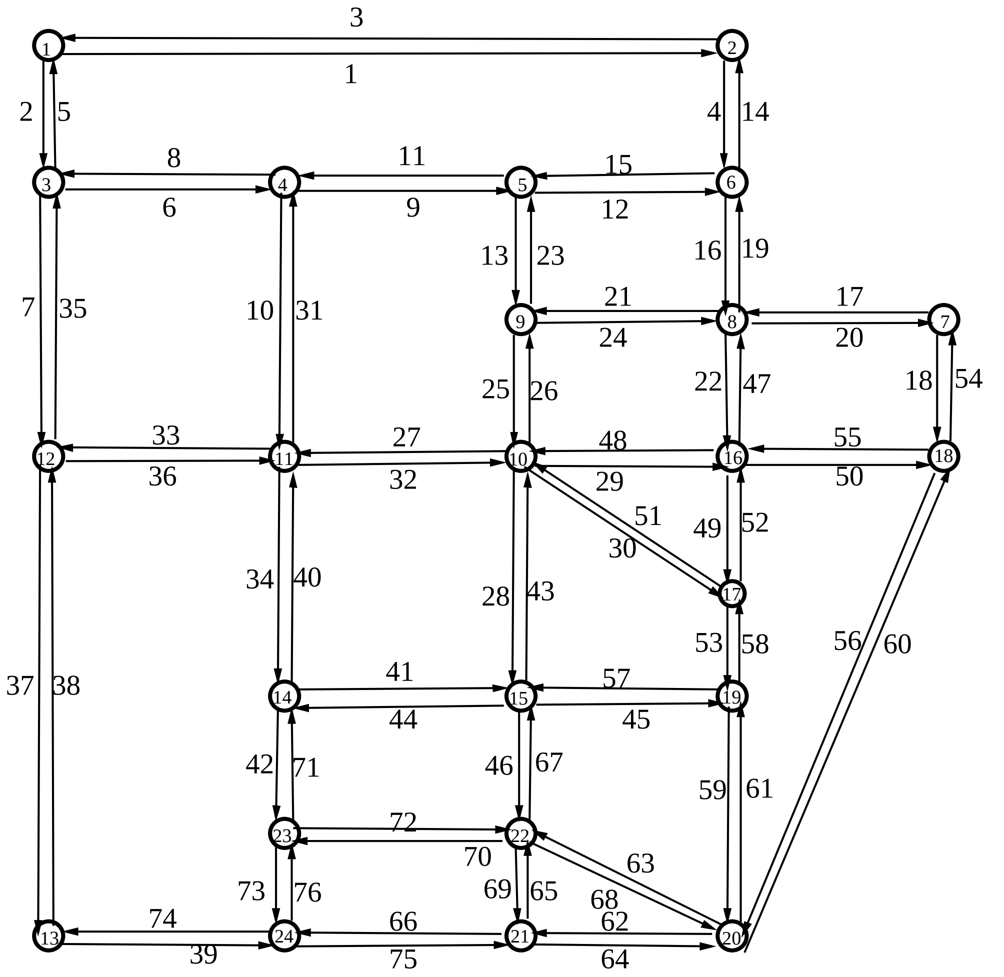
## Nguyen and Dupuis Network with random OD demand $b_\xi$ random link capacities (affecting travel time $F(\xi, \cdot)$)




13 nodes, 19 links, 25 paths connecting 4 origin-destination (OD) pairs

$1 \to 2$, $4 \to 2$, $1 \to 3$ and $4 \to 3$.

## Sioux Falls network




24 nodes, 76 links, 528 OD pairs, 1179 paths

# Wardrop's user equilibrium

- Wardrop's user equilibrium   At the equilibrium point no traveler can change his route to reduce his travel cost.
- For one scenario $\xi \in \Xi$, the static Wardrop's user equilibrium is equivalent to NCP: Find $x$, such that

$$x \geq 0, \quad F(\xi, x) \geq 0, \quad x^T F(\xi, x) = 0,$$

where $y$ : a path flow pattern, $v$ : a travel cost vector.

$$x = \begin{pmatrix} y \\ v \end{pmatrix}, \quad F(\xi, x) = \begin{pmatrix} G(\xi, y) - A^T v \\ Ay - b_\xi \end{pmatrix}.$$

VI: Find $x \in X_\xi$, such that

$$(y - x)^T G(\xi, x) \geq 0, \quad \forall y \in X_\xi = \{x | Ax = b_\xi, x \geq 0\}$$

$G$ : path travel cost function
$A$ : Origin-Destination(OD) route incidence matrix
$b$ : demand on each OD-pair

# Smoothing algorithms

- Choose a smoothing function $\tilde{\varphi}(x, \mu)$ and an algorithm for smooth problems
- Use $\tilde{\varphi}(x_k, \mu_k)$ and its gradient $\nabla\tilde{\varphi}(x_k, \mu_k)$ at each step of the algorithm
- Update the smoothing parameter $\mu_k$ at each step. The updating scheme plays a key role in convergence analysis of the smoothing method.

Challeges:

1 How to choose a smoothing function and an algorithm for the problem ?

2 How to update the smoothing parameter $\mu_k$ ?

We develop efficient smoothing projected gradient method and smoothing conjugate gradient method.
We prove global convergence of these methods to a stationary point.

# Smoothing gradient method

Step 1. Choose constants $\sigma,\ \rho \in (0,1)$, and an initial point $x^0$. Set $k = 0$.

Step 2. Compute the gradient

$$g_k = \nabla \tilde{\varphi}(x^k, \mu_k).$$

Step 3. Compute the step size $\nu_k$ by the Armijo line search, where $\nu_k = \max\{\rho^0, \rho^1, \cdots\}$ and $\rho^i$ satisfies

$$\tilde{\varphi}(x^k - \rho^i g_k, \mu_k) \leq \tilde{\varphi}(x^k, \mu_k) - \sigma \rho^i g_k^T g_k.$$

Set $x^{k+1} = x^k - \nu_k g_k$.

Step 4. If $\|\nabla \tilde{\varphi}(x^{k+1}, \mu_k)\| \geq n\mu_k$, then set $\mu_{k+1} = \mu_k$; otherwise, choose $\mu_{k+1} = \sigma \mu_k$.

Smoothing conjugate gradient method Chen-Zhou (SIIMS 2010).

# Nguyen and Dupuis Newtwork ($\beta = 0.9, \varepsilon = 3.3E3$)

| | | $x_{\rm EV}$ | $x_{\rm ERM}$ |
|---|---|---|---|
| $N = 10^3$, $\mu = 10^{-4}$ | $\text{prob}\{f(\xi, x) \le \varepsilon\}$ | 0.508 | 0.952 |
| | $E[f(\xi, x)]$ | 3.498E3 | 2.938E3 |
| | $\alpha^*$ | 7.935E3 | 3.226E3 |
| | $\text{CVaR}(x, \alpha^*)$ | 8.154E3 | 3.333E3 |
| $N = 5 * 10^3$, $\mu = 10^{-5}$ | $\text{prob}\{f(\xi, x) \le \varepsilon\}$ | 0.510 | 0.908 |
| | $E[f(\xi, x)]$ | 3.498E3 | 2.983E3 |
| | $\alpha^*$ | 7.918E3 | 3.286E3 |
| | $\text{CVaR}(x, \alpha^*)$ | 8.121E3 | 3.403E3 |
| $N = 10^4$, $\mu = 10^{-6}$ | $\text{prob}\{f(\xi, x) \le \varepsilon\}$ | 0.509 | 0.927 |
| | $E[f(\xi, x)]$ | 3.505E3 | 2.976E3 |
| | $\alpha^*$ | 7.978E3 | 3.253E3 |
| | $\text{CVaR}(x, \alpha^*)$ | 8.168E3 | 3.359E3 |

$$\alpha^*(x) \in \operatorname*{argmin}_{\alpha \in R} \text{CVaR}(x, \alpha) := \alpha + \frac{1}{1-\beta} E\{[f(\xi, x) - \alpha]_+\}.$$

# References

1. X. Chen and M. Fukushima, Expected residual minimization method for stochastic linear complementarity problems, Math. Oper. Res. 30(2005) 1022-1038.
2. X. Chen, R. J-B Wets and Y. Zhang, Stochastic variational inequalities: residual minimization smoothing/sample average approximations, SIAM J. Optim. under revision 2011.
3. X. Chen, C. Zhang and M. Fukushima, Robust solution of monotone stochastic linear complementarity problems, Math. Program. 117(2009) 51-80.
4. H. Fang, X. Chen and M. Fukushima, Stochastic R0 matrix linear complementarity problems, SIAM J. Optim. 18(2007) 482-506.
5. G. Lin, X. Chen and M. Fukushima, Solving stochastic mathematical programs with equilibrium constraints via approximation and smoothing implicit programming with penalization, Math. Program. 116(2009) 343-368.
6. C. Zhang and X. Chen, Smoothing projected gradient method and its application to stochastic linear complementarity problems, SIAM J. Optim. 20(2009) 627-649.
7. C. Zhang, X. Chen and A Sumalee, Robust Wardrop's user equilibrium assignment under stochastic demand and supply: expected residual minimization approach, Transp. Res. B 45(2011) 534-552.