Adjustable Robust Optimization Using Metamodels

İhsan Yanıkoğlu, **Dick den Hertog**, Jack P.C. Kleijnen

Tilburg University, Tilburg, The Netherlands

MASCOT-SAMO 2013 Meeting, Nice, France

1–4 July, 2013

Main idea

- Metamodels often contain many sources of **uncertainties**.
- Goal 1: find optimal solution that is **robust** against these uncertainties.
- Goal 2: find **adjustable** robust optimal solution.
- Methodology used: (Adjustable) Robust Optimization.
- We focus on uncertain input (“environmental”) factors.

Robust Optimization

- Optimization problems often contain **uncertain parameters** due to errors in estimation, implementation and measurement:

$$\min_{\mathbf{x}} \left\{ f(\mathbf{x}) \mid g_i(\mathbf{x}, \mathbf{a}) \leq 0 \quad \forall i \right\},$$

where $\mathbf{x}$ is the optimization variable, and $\mathbf{a}$ the uncertain parameter.

- The goal of **Robust Optimization** (RO) is to find solutions that are "immune" to uncertainty of parameters in a given mathematical optimization problem.

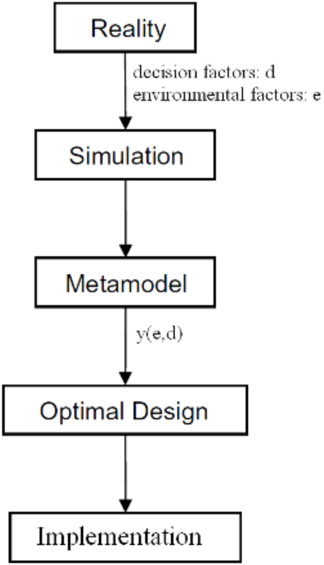
- Paradigm: use uncertainty set U for $\mathbf{a}$, and solve the robust counterpart problem:

$$\min_{\mathbf{x}} \{ f(\mathbf{x}) \mid g_i(\mathbf{x}, \mathbf{a}) \leq 0 \quad \forall \mathbf{a} \in U, \ \forall i \}.$$

Literature

- Soyster (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. (*Oper. Res.*)
- Started in the late 90s with Ben-Tal, El Ghaoui and Nemirovski.
- Bertsimas and Sim (2004). The price of robustness. (*Oper. Res.*)
- Many practical applications in logistics, finance, engineering, etc.
- Ben-Tal, El Ghaoui and Nemirovski (2009). Robust Optimization. Princeton Press.

Optimization using metamodels



Sources of uncertainties/errors

- Environmental factors;
- Simulation error;
- Metamodel error;
- Implementation error.

In this presentation we focus on **environmental** factors.

Taguchian Regression Metamodels

We focus on Taguchian Regression Metamodels:

$$y(\mathbf{e}, \mathbf{d}) = \beta_0 + \beta^T \mathbf{d} + \mathbf{d}^T B \mathbf{d} + \gamma^T \mathbf{e} + \mathbf{d}^T \Delta \mathbf{e} + \epsilon$$

- Decision factors ($\mathbf{d}$): inputs under control of users (e.g., the number of forklifts, conveyors or shipping doors in a distribution center)
- Environmental factors ($\mathbf{e}$): inputs not controlled by users (e.g., suppliers' production interruption and the quantity variability)
- Quadratic in $\mathbf{d}$, linear in $\mathbf{e}$
- Low dimension in $\mathbf{e}$ and $\mathbf{d}$

However, same methodology can be applied to other metamodels (RBF, Kriging, etc.).

Regression output predictors: Expectation and variance

Remember:

$y(\mathbf{e},\mathbf{d}) = \beta_0 + \beta^T\mathbf{d} + \mathbf{d}^T B\mathbf{d} + \gamma^T\mathbf{e} + \mathbf{d}^T\Delta\mathbf{e} + \epsilon$

$$\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \beta_0 + \beta^T\mathbf{d} + \mathbf{d}^T B\mathbf{d} + \gamma^T\mu_{\mathbf{e}} + \mathbf{d}^T\Delta\mu_{\mathbf{e}}$$

$$\mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = (\gamma^T + \mathbf{d}^T\Delta)\mathrm{Cov}(\mathbf{e})(\gamma + \Delta^T\mathbf{d})$$

Robust Parameter Design formulations in literature

Mean-Variance: $\min_{\mathbf{d}} \; \mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \;\; \text{s.t.} \;\; \mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \leq T$

Target T: $\min_{\mathbf{d}} \; \mathrm{E}_{\mathbf{e}}[(y(\mathbf{e},\mathbf{d}) - T)^2]$

Assumption 1: known mean and covariance of **e**
Assumption 2: normally distributed **e**

Goal of this research

1. Develop Robust Optimization approach:
   - that does not need these assumptions;
   - that works with pure historical data;
   - that can be used for many types of metamodels (e.g., polynomial regression, Kriging);
2. Develop Adjustable Robust Optimization approach:
   - where optimal decisions are adjustable according to environmental factor(s)' realization(s).

Limitation: method is suitable for **low** dimensions.

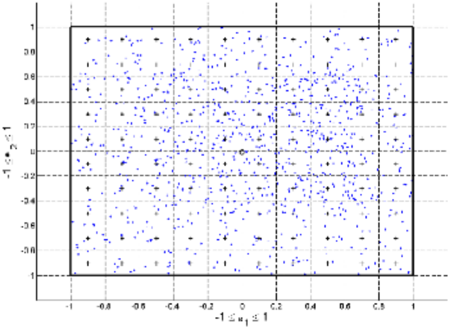
Four steps

- Step 1: Discretize w.r.t. $\mathbf{e}$;
- Step 2: Use historical data to define (ϕ-divergence) uncertainty set for $\mathbf{e}$;
- Step 3: Reformulate Robust Counterpart problem into tractable one;
- Step 4: Solve the resulting problem and analyze the robust solution.

Limitation: method is suitable for **low** dimensions.

Step 1: Discretization of $\mathbf{e}$

$$y_i(\mathbf{d}) = y(\mathbf{e}^i, \mathbf{d}) = \beta_0 + \beta^T \mathbf{d} + \mathbf{d}^T \mathbf{B} \mathbf{d} + (\gamma^T + \mathbf{d}^T \Delta)\mathbf{e}^i$$



Set of all cells: $V = \{1, \ldots, m\}$

Preliminary: ϕ-divergence

The ϕ-divergence ("distance") between two vectors $\mathbf{p} = (p_1, ..., p_m) \geq 0$ and $\mathbf{q} = (q_1, ..., q_m) \geq 0$ is

$$I_\phi(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right), \tag{1}$$

where $\phi(t)$ is convex for $t \geq 0$, $\phi(1) = 0$, and $\phi(0/0) = 0$.

Ben-Tal et al. (2012) use ϕ-divergence as uncertainty set for uncertain probability vectors.

Literature

- *Ben-Tal, Den Hertog, De Waegenaere, Melenberg and Rennen. 2012. "Robust solutions of optimization problems affected by uncertain probabilities". Management Science. Forthcoming.*

Preliminary: ϕ-divergence examples

ϕ-Divergence Examples

Divergence	$\phi(t), t > 0$	$I_\phi(p,q)$	$\phi^*(s)$
Kullback-Leibler	$t \log t$	$\sum_i p_i \log\left(\frac{p_i}{q_i}\right)$	e^{s-1}
Burg entropy	$-\log t$	$\sum_i q_i \log\left(\frac{p_i}{q_i}\right)$	$-1-\log(-s), s \leq 0$
χ^2-distance	$\frac{1}{t}(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{p_i}$	$2 - 2\sqrt{1-s}, s \leq 1$
Pearson χ^2-distance	$(t-1)^2$	$\sum_i \frac{(p_i - q_i)^2}{q_i}$	$s + s^2/4, s \geq -2$ $-1, s < -2$
Hellinger distance	$(1-\sqrt{t})^2$	$\sum_i (\sqrt{p_i} - \sqrt{q_i})^2$	$\frac{s}{1-s}, s \leq 1$

Preliminary: Convex conjugate function

Convex conjugate function defined as follows:

$$\phi^* (s) := \sup_{t \geq 0} \{st - \phi (t)\} .$$

We work with ϕ-divergence distances for which the closed-form conjugates are available.

Preliminary (cont'd)

Remember: $I_\phi(\mathbf{p},\mathbf{q}) := \sum_{i=1}^{m} q_i \phi\left(\frac{p_i}{q_i}\right)$ (1)

- We consider $\mathbf{p} = (p_1, ..., p_m)$ in (1) as the unknown true probability vector of uncertain parameter $\mathbf{e} \in \mathbb{R}^L$.

- $\mathbf{q} = (q_1, ..., q_m)$ are observed frequencies in the historical data.

Remember: We assume there are m cells, $V := \{1, \ldots, m\}$.

Step 2: ϕ-divergence uncertainty set

Using the chi-squared test statistic, an approximate $(1-\alpha)$-confidence set (U) for p is

$$U := \left\{ \mathbf{p} \in \mathbb{R}^m : \mathbf{p} \geq 0, \mathbf{p}^T \mathbf{1} = 1, I_\phi(\mathbf{p}, \mathbf{q}) \leq \rho \right\},$$

where

$$\rho = \rho_\phi(N, m-1, \alpha) := \frac{\phi''(1)}{2N} \chi^2_{m-1,1-\alpha},$$

N is the sample size.

Step 3: Robust reformulation of the model

The robust reformulation of the expectation-variance (E&V) model

$$\min_{\mathbf{d}} \ \mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \ \text{ s.t. } \ \mathrm{Var}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \leq T$$

is the following *semi-infinite* optimization problem:

$$\min_{\mathbf{d}} \max_{\mathbf{p}\in U} \hat{\mathrm{E}}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]$$
$$\text{s.t. } \ \widehat{\mathrm{Var}}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \leq T \quad \forall \mathbf{p} \in U.$$

Remember: $\mathrm{E}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \beta_0 + \beta^{^T}\mathbf{d} + \mathbf{d}^{^T}B\mathbf{d} + \gamma^{^T}\mu_{\mathbf{e}} + \mathbf{d}^{^T}\Delta\mu_{\mathbf{e}},$

$$\hat{\mathrm{E}}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \sum_{i\in V}[\beta_0 + \beta^{^T}\mathbf{d} + \mathbf{d}^{^T}\mathbf{B}\mathbf{d} + (\gamma^{^T} + \mathbf{d}^{^T}\Delta)\mathbf{e}^i]$$

$$\widehat{\mathrm{Var}}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] = \sum_{i\in V}\psi_i(\mathbf{d})^2 p_i - [\sum_{i\in V}\psi_i(\mathbf{d})p_i]^2$$

Step 3: Robust counterpart of E&V model (cont'd)

$$\textbf{(RC)} \ \min_{\mathbf{d}} \ \max_{\mathbf{p}\in U} \sum_{i\in V} [\beta_0 + \beta^T \mathbf{d} + \mathbf{d}^T \mathbf{B}\mathbf{d} + (\gamma^T + \mathbf{d}^T \Delta)\mathbf{e}^i] p_i$$

$$\text{s.t.} \ \sum_{i\in V} \psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V} \psi_i(\mathbf{d}) p_i\right]^2 \leq T \quad \forall \mathbf{p} \in U,$$

Remember: $U := \left\{\mathbf{p} \in \mathbb{R}^m : \mathbf{p} \geq 0, \mathbf{p}^T\mathbf{1} = 1, I_\phi(\mathbf{p}, \mathbf{q}) \leq \rho\right\},$

$\psi_i(\mathbf{d}) := (\gamma^T + \mathbf{d}^T \Delta)\mathbf{e}^i$

Tractable robust counterpart of E&V model

Theorem

The vector $\mathbf{d}$ *solves (RC) if and only if* $\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\eta}$, *and* z *solve the following problem:*

$$\min_{\mathbf{d},\boldsymbol{\lambda},\boldsymbol{\eta},z} \; \beta_{\mathbf{0}} + \boldsymbol{\beta}^T \mathbf{d} + \mathbf{d}^T \mathbf{B} \mathbf{d} + \lambda_1 + \rho\eta_1 + \eta_1 \sum_{i \in V} q_i \phi^* \left(\frac{\psi_i(\mathbf{d}) - \lambda_1}{\eta_1} \right)$$

$$\text{s.t.} \;\; \lambda_2 + \rho\eta_2 + \eta_2 \sum_{i \in V} q_i \phi^* \left(\frac{(\psi_i(\mathbf{d}) + z)^2 - \lambda_2}{\eta_2} \right) \leq T$$

$$\eta_1, \eta_2 \geq 0$$

Literature

- *Ben-Tal, A., D. den Hertog, J.-P. Vial. 2012. "Deriving robust counterparts of nonlinear uncertain inequalities". CentER Discussion Paper No. 2012-053, Tilburg University.*
- *Yanıkoğlu, İ., D. den Hertog. 2012. "Safe approximations of ambiguous chance constraints using historical data". INFORMS Journal on Computing. Forthcoming.*

Example: Color television images

The response model of the quality of transmitted signals is

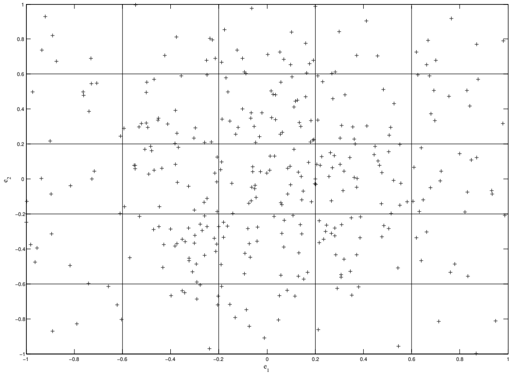
$$\begin{aligned} \hat{y}(\mathbf{d}, \mathbf{e}) &= 33.389 - 4.175d_1 + 3.748d_2 + 3.348d_1d_2 - 2.328d_1^2 \\ &- 1.867d_2^2 - 4.076e_1 + 2.985e_2 - 2.324d_1e_1 \\ &+ 1.932d_1e_2 + 3.268d_2e_1 - 2.073d_2e_2. \end{aligned}$$

- control: number of tabs in a filter (d_1), sampling frequency (d_2)
- environmental: number of bits in an image (e_1), voltage applied (e_2)

Source

- *Myers, R. H., D. C. Montgomery, C. M. Anderson-Cook. 2009. Response surface methodology: process and product optimization using designed experiments, 3rd ed.*

Step 1: Discretize w.r.t. **e**



Historical data on **e**

Step 2: Define ϕ-divergence uncertainty set

Uncertainty Set:

$$U := \left\{ \mathbf{p} \in \mathbb{R}^{25} : \mathbf{p} \geq 0, \mathbf{p}^T \mathbf{1} = 1, I_\phi(\mathbf{p}, \mathbf{q}) \leq \rho \right\}$$

Nominal Problem:

$$\max_{\mathbf{d}} \sum_{i \in V} [\beta_0 + \beta^T \mathbf{d} + \mathbf{d}^T \mathbf{B} \mathbf{d} + (\gamma^T + \mathbf{d}^T \Delta) \mathbf{e}^i] q_i$$

$$\text{s.t.} \quad \sum_{i \in V} \psi_i(\mathbf{d})^2 q_i - \left[\sum_{i \in V} \psi_i(\mathbf{d}) q_i \right]^2 \leq T$$

Step 3: Reformulate robust counterpart

$$\max_{\mathbf{d}} \ \mathbb{E}_{\mathbf{e}}[\hat{y}(\mathbf{d},\mathbf{e})]$$

$$\mathrm{Var}_{\mathbf{e}}[\hat{y}(\mathbf{d},\mathbf{e})] \leq T.$$

$$\underline{\text{Remember}} : \max_{\mathbf{d}} \min_{\mathbf{p}\in U} \hat{\mathrm{E}}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})]$$

$$\text{s.t. } \widehat{\mathrm{Var}}_{\mathbf{e}}[y(\mathbf{e},\mathbf{d})] \leq T \quad \forall \mathbf{p} \in U.$$

$$\underline{\text{Remember}} : \max_{\mathbf{d}} \ \min_{\mathbf{p}\in U} \sum_{i\in V} [\beta_0 + \beta^T \mathbf{d} + \mathbf{d}^T \mathbf{B}\mathbf{d} + (\gamma^T + \mathbf{d}^T \Delta)\mathbf{e}^i] p_i$$

$$\text{s.t. } \sum_{i\in V} \psi_i(\mathbf{d})^2 p_i - \left[\sum_{i\in V} \psi_i(\mathbf{d}) p_i\right]^2 \leq T \quad \forall \mathbf{p} \in U$$

Step 4: Solve and analyze the robust solution

Worst-Case Analysis

T	Robust		Nominal	
	y^*	Var.(%)	W.-C.($\mathbb{E}$)	W.-C.(Var)(%)
0.1	35.27	0.1(0%)	35.279	0.188(88%)
0.2	35.28	0.149(-25%)	35.261	0.375(87%)
0.3	35.28	0.149(-50%)	35.233	0.559(86%)
0.4	35.28	0.149(-62%)	35.204	0.715(79%)

Step 4: Solve and analyze the robust solution

Average Analysis

T	Robust		Nominal	
	Avg($\mathbb{E}$)	Avg(Var)	Avg($\mathbb{E}$)	Avg(Var)
0.1	35.3776	0.0562	35.4168	0.1051
0.2	35.4023	0.0836	35.4572	0.2095
0.3	35.4023	0.0836	35.4730	0.3138
0.4	35.4023	0.0836	35.4765	0.4012

Adjustable Robust Optimization (ARO)

Two types of decisions in ARO:

- 'here and now': decisions on $\mathbf{d}$ must be made before $\mathbf{e}$ is realized;
- 'wait and see': decisions on $\mathbf{d}$ are adjusted after observing (some of) the actual values of $\mathbf{e}$.

Examples:

- design of distribution center;
- design of multi component product.

To model this situation, we reformulate d_j as a function of $\mathbf{e}$:

$$d_j = d_j(\mathbf{e}).$$

Advantages of ARO

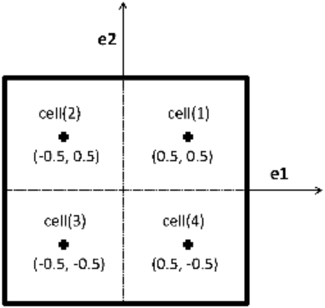
- Information on e is used in "wait-and-see" decision.
- The optimal solution of ARC is less conservative than that of the non-adjustable RC.
- The complexity of the RC and ARC are the same, although extra optimization variables are added.
- It has many applications in real-life.

Illustrative example

General Problem:

$$\min_{\mathbf{d}} \ \mathrm{E}_{\mathbf{e}}\left[(1 + 5d_1 + 5d_2 + e_1 - e_2)^2 + (1 + 5d_1 + 10d_2 + e_1 + e_2)^2\right].$$

Discretize w.r.t. e:



Illustrative example

Nominal Problem:

$$\min_{\mathbf{d}} \sum_{i=1}^{4} \left[\left(1 + 5d_1 + 5d_2 + e_1^i - e_2^i\right)^2 + \left(1 + 10d_2 + 5d_1 + e_1^i + e_2^i\right)^2 \right] q_i .$$

- $\mathbf{q} = \{0.4, 0.3, 0.2, 0.1\}$

The optimal solution is attained at (d_1, d_2)=**(-0.08, -0.08)**

Illustrative example

Robust Counterpart:

$$\min_{\mathbf{d}} \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{4} \left[\left(1 + 5d_1 + 5d_2 + e_1^i - e_2^i\right)^2 + \left(1 + 5d_1 + 10d_2 + e_1^i + e_2^i\right)^2 \right] p_i,$$

- $\mathcal{P} := \left\{ \mathbf{p} = (p_1, p_2, p_3, p_4) \in \mathbb{R}^4 \middle| p_1 + p_2 + p_3 + p_4 = 1, \sum_{i=1}^{4} \frac{(p_i - q_i)^2}{p_i} \leq 0.5, \mathbf{p} \geq 0 \right\}$

The robust optimal solution is **(-0.2, 0)** with objective value **1**

Remark: The nominal solution **(-0.08, -0.08)** has objective value **1.2** in the worstcase.

Illustrative example: Adjustable robust optimization

Linear Decision Rule: The controllable factors **d** are linearly adjustable on the observed values of the environmental factors **e**:

$$d_j = d_j(\mathbf{e}) := x_{j0} + x_{j1}e_1 + x_{j2}e_2 \quad \forall j \in \{1, 2\}.$$

Illustrative example: Adjustable robust optimization

Consequently, the ARC of the illustrative example is

$$\min_{\mathbf{x}} \max_{\mathbf{p}\in\mathcal{P}} \sum_{i=1}^{4} \Bigg[\Big(1 + 5d_1(\mathbf{e^i}) + 5d_2(\mathbf{e^i}) + e_1^i - e_2^i\Big)^2 + \Big(1 + 5d_1(\mathbf{e^i}) + 10d_2(\mathbf{e^i}) + e_1^i + e_2^i\Big)^2 \Bigg] p_i.$$

The optimal solution for fully adjustable d_1 and d_2 is

$$d_1(\mathbf{e}) = -1/5 - 1/5e_1 + 3/5e_2$$
$$d_2(\mathbf{e}) = -2/5e_2.$$

The optimal objective value is **0**.

Remark: It was **1** for the RC, and **1.2** for the nominal (worstcase).

Illustrative example: Linear decision rules

- **na**: non-adjustable, $d_j = x_{j0}$
- **e1**: $d_j = x_{j0} + x_{j1}e_1$
- **e2**: $d_j = x_{j0} + x_{j2}e_2$
- **e[1,2]**: $d_j = x_{j0} + x_{j1}e_1 + x_{j2}e_2$

Illustrative example: Linear decision rules

Linear Decision Rules

D.R.		$d_1(.)$	$d_2(.)$	Obj.
d_1	d_2	(x_{10}, x_{11}, x_{12})	(x_{20}, x_{21}, x_{22})	
na	na	(-0.2,−,−)	(0,−,−)	1
na	e1	(-0.189,−,−)	(-0.009,-0.103,−)	0.662
na	e2	(-0.2,−,−)	(0,−,0)	1
e1	na	(-0.2,-0.2,−)	(0,−,−)	0.5
e2	na	(-0.2,−,0)	(0,−,−)	1
na	e[1,2]	(-0.186,−,−)	(-0.008,-0.12,-0.04)	0.621
e[1,2]	na	(-0.2,-0.2,0)	(0,−,−)	0.5
e1	e1	(-0.2,-0.2,−)	(0,−,−)	0.5

Illustrative example (cont'd)

Linear Decision Rules

D.R.		$d_1(.)$	$d_2(.)$	Obj.
d_1	d_2	(x_{10}, x_{11}, x_{12})	(x_{20}, x_{21}, x_{22})	
e1	e2	(-0.2,-0.2,0)	(0,$-$,-0.04)	0.45
e2	e1	(-0.21,$-$,0.04)	(-0.001,-0.09,$-$)	0.68
e2	e2	(-0.2,$-$,0.6)	(0,$-$,-0.4)	0.5
e[1,2]	e1	(-0.2,-0.2,0)	(0,$-$, $-$)	0.5
e[1,2]	e2	(-0.2,-0,2,0.6)	($-$, $-$,0.4)	0
e1,	e[1,2]	(-0.2,-0.2,$-$)	(0,0,-0.04)	0.45
e2,	e[1,2]	(-0.2,$-$,0.6)	(0,-0.12,-0.4)	0.05
e[1,2]	e[1,2]	(-0.2,-0.2,0.6)	(0,0,-0.4)	0

Conclusions

- We used Robust Optimization methodology to find robust solution.
- Metamodels often contain uncertain input (environmental) factors.
- Our approach directly uses historical data via ϕ-divergence.
- It has no assumptions on type of distribution and the value of its parameter(s).
- It leads to tractable robust counterpart problems.
- Adjustable robust optimization can be used in the "wait-and-see" situation.
- **Limitation:** We may require a lot of data observations, especially when the number of uncertain parameters is high.