

SLP based method: (further details)

# Solution procedures

$$\min_{x \in \mathcal{N}^a} \mathbb{E}\{f(\xi, x(\xi))\} = \min_{x^1 \in \mathbb{R}^{n_1}} f_1(x^1) + EQ_1(x^1)$$

$$EQ_1(\xi; x^1) = \mathbb{E}\left\{\inf_{x^2 \in \mathbb{R}^{n_2}} f_2(\xi; x^1, x^2) + EQ_2(\xi; x^1, x^2) \middle| \mathcal{A}_1\right\}$$

$$EQ_2(\xi; x^1, x^2(\xi)) = \mathbb{E}\left\{\inf_{x^3 \in \mathbb{R}^{n_3}} f_3(\xi; x^1, x^2(\xi), x^3) \middle| \mathcal{A}_2\right\}$$

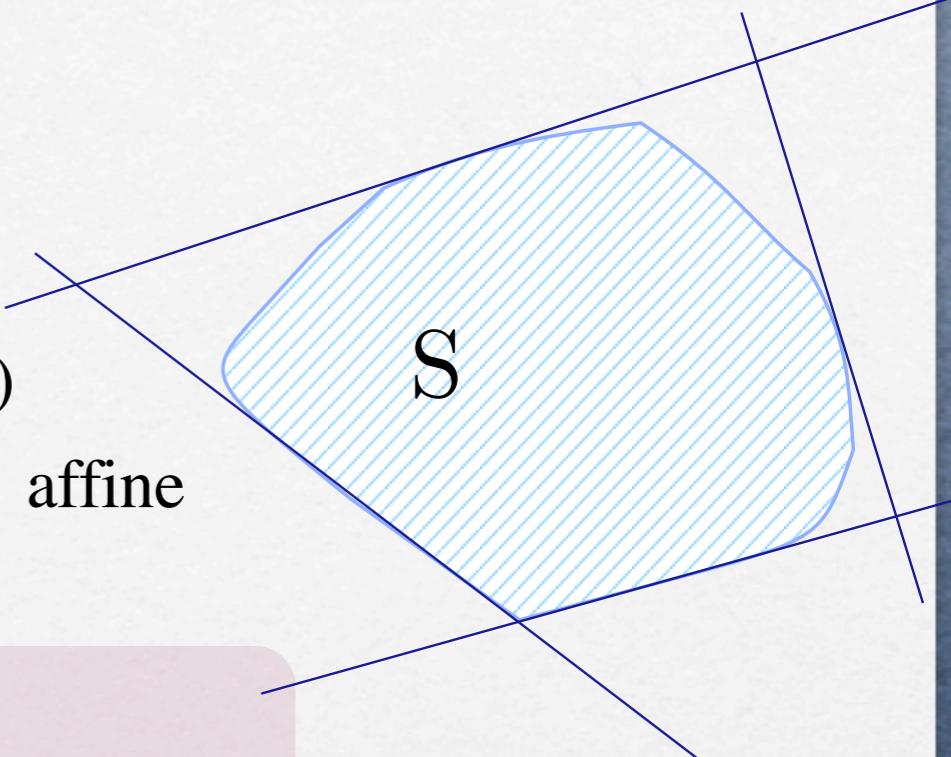
deterministic optimization problem with  $f$  convex random lsc function

# Sequential I.p. strategy

$\min f_0(x), \quad x \in X \subset \mathbb{R}^n, \quad f_0$  linear (not essential)

$f_i(x) \leq 0, \quad i = 1, \dots, s, \quad f_i(s) = 0, \quad i = s + 1, \dots, m$  (affine)

in the  $s + 1$  first constraints:  $f_i(x) = \sup_{t \in T} f_{i,t}(x), \quad f_i \geq f_{i,t}$  affine



0.  $v = 0$ , pick polytope (box)  $K^0 \ni x^{opt}$

1.  $x^v \in \arg \min f_0$  on  $K^v$ , set  $i_v : f_{i_v}(x^v) = \max_{1 \leq i \leq s} f_i(x^v)$

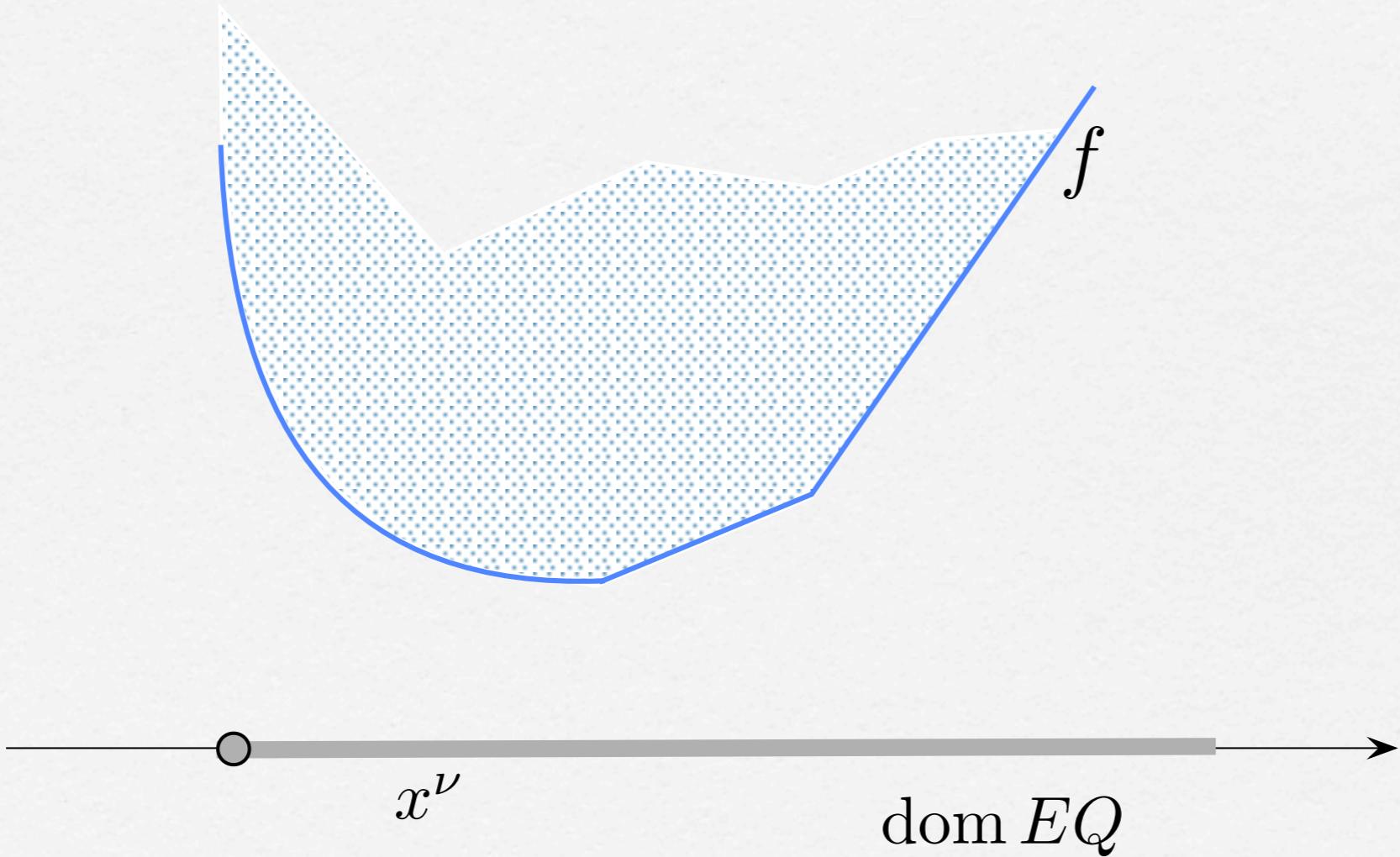
if  $f_{i_v}(x^v) \leq 0$ ,  $x^v$  optimal, otherwise go to 2.

2. return to 1. with  $K^{v+1} = K^v \cap \left\{ \langle \nabla f_{i_v}(x^v), x - x^v \rangle + f_{i_v}(x^v) \leq 0 \right\}$

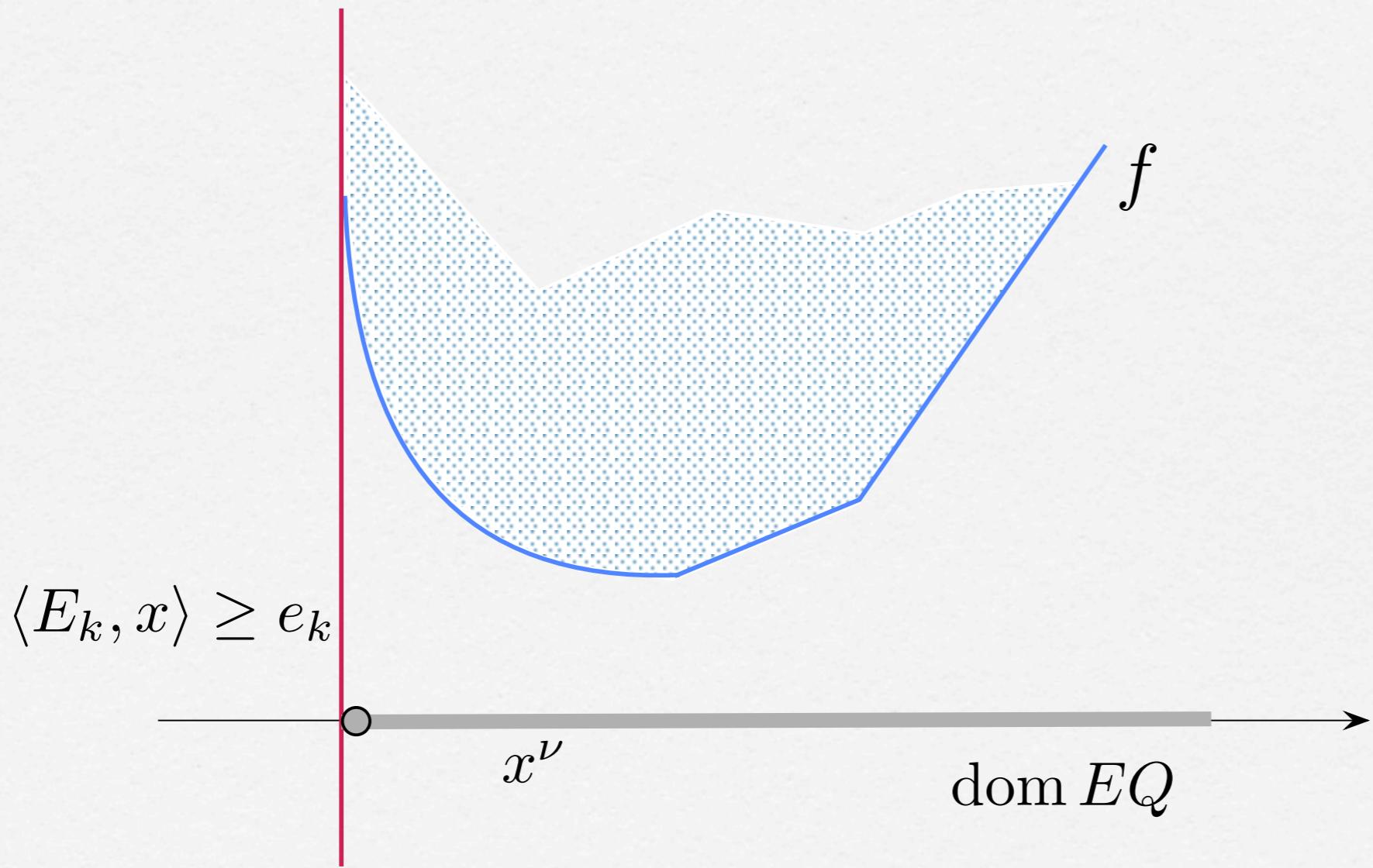
when  $f_0$  is not linear (but convex):  $\min \theta$  such that  $f_0(x) - \theta \leq 0$

convergence: finite # of steps or iterates cluster to optimal sol'n

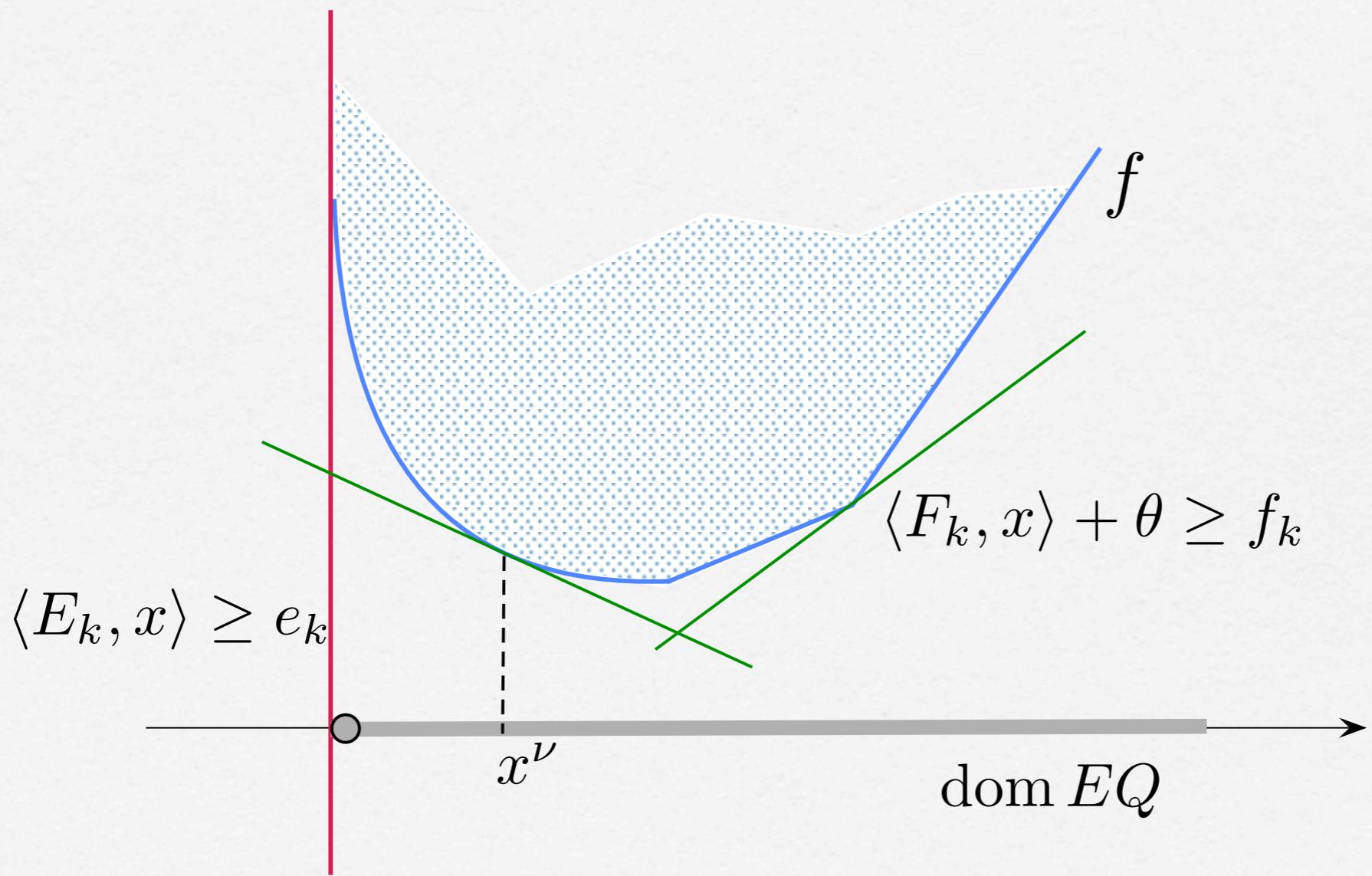
# Generating cutting hyperplanes



# Generating cutting hyperplanes



# Generating cutting hyperplanes



# SLP for Stochastic Programs

$$\min f_1(x) + EQ_1(x) \text{ s.t. } Ax = b, x \geq 0 \quad (x = x^1)$$

$$EQ_1(x) = \sum_{l=1}^L p_l Q_1(\xi^l, x) \quad L \text{ large}$$

$$Q_1(\xi^l, x) = \inf_{x^2 \in X_2} \left\{ f_2(\xi^l; x, x^2) + (EQ_2(\dots)) \right\}$$

$$\text{dom } EQ_1 = \bigcap_{l=1}^L \text{dom } Q_1(\xi^l, \cdot) = \bigcap_{l=1}^L \left\{ x \mid \exists x^2 \in X_2, f_2(\xi^l; x, x^2) < \infty \right\}$$

0.  $v = r = s = 0$

1.  $v = v + 1$ , solve:  $\min f_1(x) + \theta, Ax = b, x \geq 0$  such that

(feasibility cuts)  $\langle E_k, x \rangle \geq e_k, k = 1 \rightarrow r$

(optimality cuts)  $\langle F_k, x \rangle + \theta \geq f_k, k = 1 \rightarrow s$

2. generate feasibility cuts: check if  $x \in \text{dom } EQ_1$ .

No:  $E_k$  separates  $x$  from  $\text{dom } EQ_1$ , go to 1. Yes, go to 3.

3. generate optimality cuts:  $F_k \in \partial EQ_1(x^k)$ , go to 1.

# Cut Generation: Fixed Recourse

(l.p.)-solution:  $(x^\nu, \theta^\nu)$

$x^\nu$  feasible?

- $\forall \xi \in \Xi : z_\xi = \operatorname{argmax}_z \left\{ \langle d_\xi - T_\xi x^\nu, z \rangle \mid W^\top z \leq 0, -1 \leq z_j \leq 1 \right\}$   
if  $\eta_\xi = \langle d_\xi - T_\xi x^\nu, z \rangle = 0$ ,  $x^\nu$  feasible  
for some  $\xi, \eta_\xi > 0$ , then  $E_{k+1} = (T_\xi)^\top z_\xi$ ,  $e_{k+1} = \langle d_\xi, z_\xi \rangle$

$x^\nu$  optimal?

$$\forall \xi \in \Xi : v_\xi = \operatorname{argmax}_v \left\{ \langle d_\xi - T_\xi x^\nu, v \rangle \mid W^\top z \leq q_\xi \right\}$$

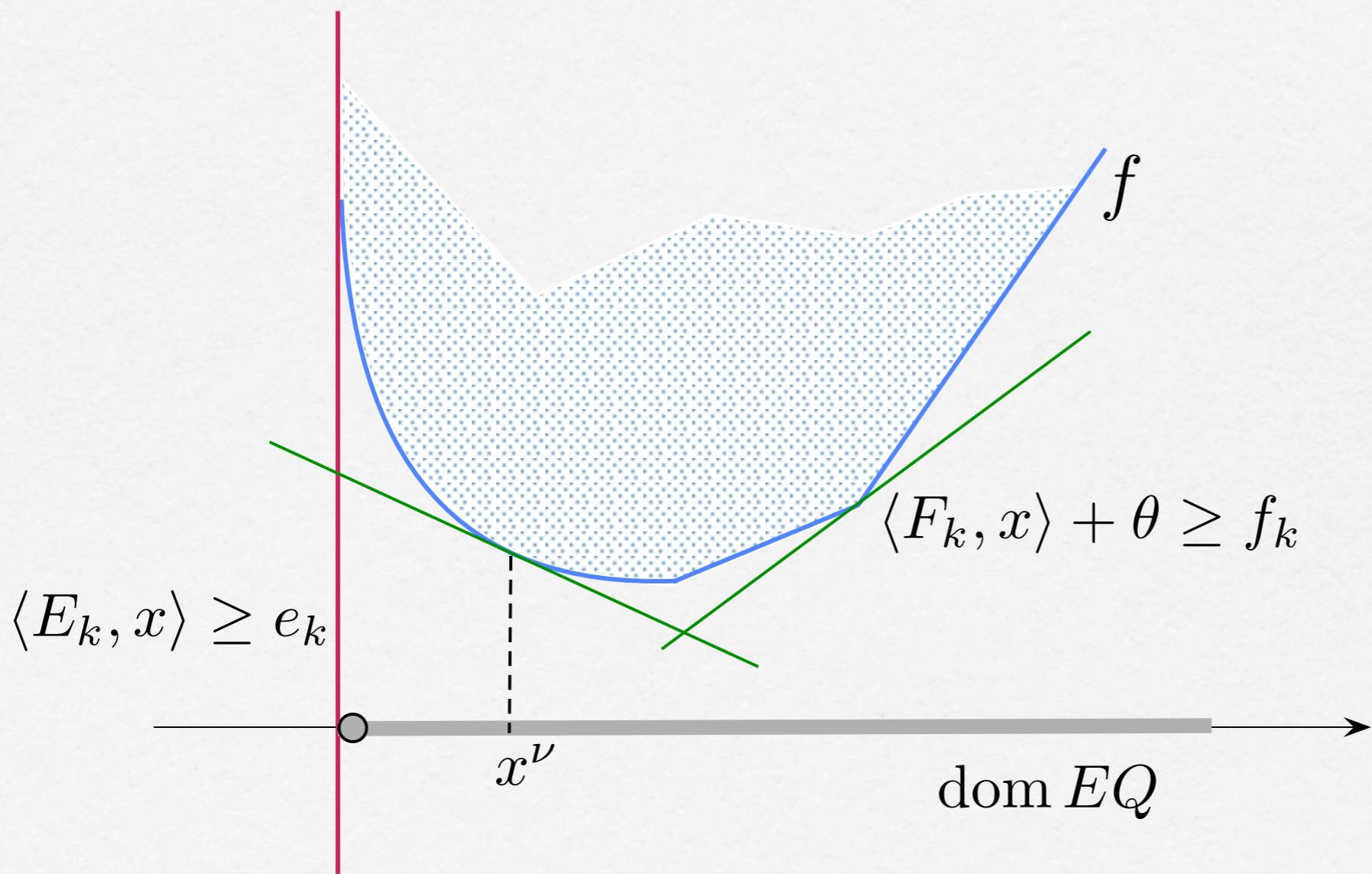
if infeasible for some  $\xi \implies$  unbounded problem

otherwise  $F_{k+1} = \mathbb{E}\{Tv\}$ ,  $f_{k+1} = \mathbb{E}\{\langle d, v \rangle\}$

if  $\theta^\nu \geq f_k + 1 - \langle F_{k+1}, x^\nu \rangle \implies x^\nu$  optimal

add optimality cut

# Generating cutting hyperplanes



# Aggregation Principle in Stochastic Optimization

Roger J-B Wets  
Mathematics, University of California, Davis

# Some Examples:

## 1. Stochastic Programming (recourse model)

$$f(\xi, x) = \begin{cases} f_{01}(x) + Q(\xi, x) & \text{if } x \in C_1 \\ \infty & \text{otherwise} \end{cases}$$

$$Q(\xi, x) = \inf_y \{ f_{02}(\xi, y) \mid y \in C_2(\xi, x) \}$$

network capacity  
expansion, e.g.

$$\min E f(x) = \mathbb{E}\{f(\xi, x)\},$$

$$\text{SAA-problem: } \min f^\nu(\vec{\xi}^\nu, x) = \frac{1}{\nu} \sum_{l=1}^\nu f(\xi^l, x)$$

## 2. Statistical Estimation (fusion of hard & soft information)

$$L(\xi, h) = \begin{cases} -\ln h(\xi) & \text{if } h \geq 0, \int h = 1, h \in A^{\text{soft}} \subset E \\ \infty & \text{otherwise} \end{cases}$$

$$EL(h) = \mathbb{E}\{L(\xi, h)\}, h^{\text{true}} = \operatorname{argmin}_E \mathbb{E}\{L(\xi, h)\}$$

$$\text{estimate: } h^\nu \in \operatorname{argmin}_E \mathbb{E}^\nu\{L(\xi, h)\} = \frac{1}{\nu} \sum_{l=1}^\nu L(\xi^l, h)$$

$A^{\text{soft}}$  : constraints on support, moments, shape, smoothness, ...

# Pricing financial instruments

3. A contingent claim:

environment process:  $\{\xi^t \in \mathbb{R}^d\}_{t=0}^T$

history:  $\vec{\xi}^t$ ,  $\vec{\xi} = \vec{\xi}^T$ , price process:  $S^t(\vec{\xi}) \in \mathbb{R}^n$ ; numéraire (risk-free):  $S_1^t \equiv 1$

claims:  $\left\{G^t(\vec{\xi})\right\}_{t=1}^T$ ;  $i$ -strategy:  $\left\{X^t(\vec{\xi})\right\}_{t=0}^T$ ; value @  $t$ :  $\langle S^t(\vec{\xi}), X^t(\vec{\xi}) \rangle$

*Instruments:* T-bonds, options, swaps, insurance contracts, mortgages, ...

# Pricing financial instruments

3. A contingent claim: environment process:  $\{\xi^t \in \mathbb{R}^d\}_{t=0}^T$   
history:  $\vec{\xi}^t$ ,  $\vec{\xi} = \vec{\xi}^T$ , price process:  $S^t(\vec{\xi}) \in \mathbb{R}^n$ ; numéraire (risk-free):  $S_1^t \equiv 1$

claims:  $\left\{G^t(\vec{\xi})\right\}_{t=1}^T$ ;  $i$ -strategy:  $\left\{X^t(\vec{\xi})\right\}_{t=0}^T$ ; value @  $t$ :  $\langle S^t(\vec{\xi}), X^t(\vec{\xi}) \rangle$

*Instruments:* T-bonds, options, swaps, insurance contracts, mortgages, ...

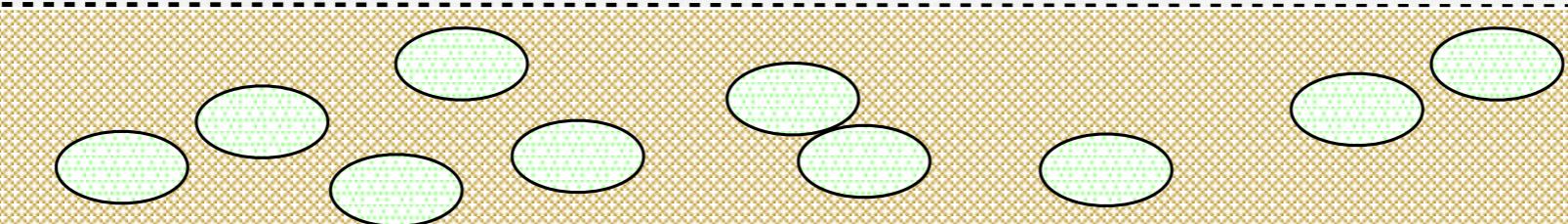
$\max \mathbb{E}\{\langle S^T, X^T \rangle\}$  such that  $\langle S^t, X^t \rangle \leq G^t + \langle S^t, X^{t-1} \rangle$ ,  $t = 1 \rightarrow T$

$$\langle S^0, X^0 \rangle \leq G^0, \langle S^T, X^T \rangle \leq G^T \text{ a.s.}$$

feasible if  $G^0 + \dots + G^T \geq 0 \quad \forall \xi$ ; arbitrage  $\Rightarrow$  unbounded

$\text{prob}[\xi = \vec{\xi}] = p_{\vec{\xi}}$  (finite sample?):  $\max \sum_{\xi \in \Xi} p_{\vec{\xi}} \langle S^T(\vec{\xi}), X^T(\vec{\xi}) \rangle \dots$

# 4. Stochastic homogenization



$$-\nabla \cdot (a(\xi, x) \nabla u(\xi, x)) = h(x) \text{ for } x \in \Omega, \quad u(\xi, x) = 0 \text{ on bdry } \Omega$$

Variational formulation:  $\forall \xi, \quad g(\xi, u) := \frac{1}{2} \int_{\Omega} a(\xi, x) |\nabla u|^2 dx - \langle h, u \rangle$   
 find  $u(\xi, x) \in \operatorname{argmin}_{u \in H_0^1(\Omega)} g(\xi, u), \quad g(\xi, \cdot) : L^2 \rightarrow (-\infty, \infty], \text{ convex}$

$\mathbb{E}\{u(\xi, x)\} \in \operatorname{argmin}_{u \in H_0^1(\Omega)} G(u)$  where  $\operatorname{epi} G = \mathbb{E}\{\operatorname{epi} g(\xi, \cdot)\}$   
 $G(u) = \inf_z \{ \mathbb{E}\{g(\xi, z(\xi)) \mid \mathbb{E}\{z(\xi)\} = u\}$   
 $G^* = \mathbb{E}\{g^*(\xi, \cdot)\}, \quad g^*(\xi, v) = \sup_u \{ \langle v, u \rangle - g(\xi, u) \}, \text{ conjugate fcn}$   
 $\xi^1, \xi^2, \dots$  stationary, use Ergodic Theorem for random lsc functions

$$G = g^{\text{hom}} = (\operatorname{epi}_w\text{-}\lim_{\nu} \frac{1}{\nu} \sum_{l=1}^{\nu} g^*(\xi^l, \cdot))^* \implies \text{values of } a^{\text{hom}}(x)$$

# Interchanging: $\mathbb{E}$ & $\min$

Evident: with  $E = \{x : \Xi \rightarrow \mathbb{R}^N \mid \text{measurable, ...}\}$

$$\min \mathbb{E} \{f(\xi, x(\xi)) \mid x \in E\} = \mathbb{E} \{ \min f(\xi, x) \mid x \in \mathbb{R}^N \}$$

when  $\exists x(\cdot) \in E$  such that  $P\text{-a.s. } x(\xi) \in \operatorname{argmin} f(\xi, \cdot)$

$x$  is measurable, ...

But our problem is:  $\min \mathbb{E}\{f(\xi, x)\}$ , equivalently,

$$\min Ef(x) = \mathbb{E}\{f(\xi, x(\xi))\}$$

such that  $x(\xi) = \mathbb{E}\{x(\xi)\}$   $P\text{-a.s.}$

$x$  can not depend on ‘anticipated’ (future) information

# Dynamic Information Process

So far,  $x$  mostly restricted to  $\{\emptyset, \Xi\}$ -measurable, i.e., constant on  $\Xi$

Generally, as  $t \nearrow T$  (possibly  $\infty$ ) additional information is acquired

$$\mathcal{A}_0 = \{\emptyset, \Xi\} \subset \mathcal{A}_1 \subset \dots \subset \mathcal{A}_T = \mathcal{A}, \text{ a filtration}$$

with  $x_t$  decision @ time  $t$  depend on available information, i.e.  $\mathcal{A}_t$ -measurable

## Reformulation

Let  $x(\xi) = (x_0(\xi), x_1(\xi), \dots, x_T(\xi)) : \Xi \rightarrow \mathbb{R}^N$ ,  $N = \sum_{t=0}^T n_t$   
 $\mathcal{N}_a = \{x \in E \mid x_t \text{ } \mathcal{A}_t\text{-measurable, } t = 0, \dots, T\}$

find  $x \in \mathcal{N}_a$  such that  $Ef(x) = \mathbb{E}\{f(\xi, x(\xi))\}$  is minimized

**Nonanticipativity constraints:**  $x \in \mathcal{N}_a$  (linear subspace)

# Here-&-Now vs. Wait-&-See

- ◆ Basic Process: decision  $\rightarrow$  observation  $\rightarrow$  decision

$$x^1 \rightarrow \xi \rightarrow x_\xi^2$$

- ◆ Here-&now problem!  $x^1$   
not all contingencies available at time 0  
can't depend on  $\xi$ !

- ◆ Wait-&-see problem  
implicitly all contingencies available at time 0  
choose  $(x_\xi^1, x_\xi^2)$  after observing  $\xi$

- ◆ incomplete information to anticipative information ?

# Stochastic Optimization: Fundamental Theorem

# Stochastic Optimization: Fundamental Theorem

A here-and-now problem can be “reduced” to a wait-and-see problem by introducing the

appropriate ‘information’ costs  
(price of non-anticipativity)

# Price of Nonanticipativity

**Here-&-now**

$$\min \mathbb{E}\{f(\xi, x^1, x_\xi^2)\}$$

$$x^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x^1), \forall \xi.$$

# Price of Nonanticipativity

## Here-&-now

$$\min \mathbb{E}\left\{f(\xi, x^1, x_\xi^2)\right\}$$

$$x^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x^1), \forall \xi.$$

## Explicit non-anticipativity

$$\min \mathbb{E}\left\{f(\xi, x_\xi^1, x_\xi^2)\right\}$$

$$x_\xi^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x_\xi^1), \forall \xi.$$

# Price of Nonanticipativity

## Here-&-now

$$\min \mathbb{E}\{f(\xi, x^1, x_\xi^2)\}$$

$$x^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x^1), \forall \xi.$$

## Explicit non-anticipativity

$$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2)\}$$

$$x_\xi^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x_\xi^1), \forall \xi.$$

$$x_\xi^1 = \mathbb{E}\{x_\xi^1\} \quad \forall \xi$$

# Price of Nonanticipativity

## Here-&-now

$$\min \mathbb{E}\{f(\xi, x^1, x_\xi^2)\}$$

$$x^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x^1), \forall \xi.$$

## Explicit non-anticipativity

$$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2)\}$$

$$x_\xi^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x_\xi^1), \forall \xi.$$


$$x_\xi^1 = \mathbb{E}\{x_\xi^1\} \quad \forall \xi$$

$w_\xi \perp$  subspace of constant fcns

$$\Rightarrow \mathbb{E}\{w_\xi\} = 0$$

# Price of Nonanticipativity

## Here-&-now

$$\min \mathbb{E}\{f(\xi, x^1, x_\xi^2)\}$$

$$x^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x^1), \forall \xi.$$

## Explicit non-anticipativity

$$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2)\}$$

$$x_\xi^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x_\xi^1), \forall \xi.$$


$$x_\xi^1 = \mathbb{E}\{x_\xi^1\} \quad \forall \xi$$

*w<sub>ξ</sub>* ⊥ subspace of constant fcns  
*multipliers*  $\Rightarrow \mathbb{E}\{w_\xi\} = 0$

$$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2) - \langle w_\xi, x_\xi^1 \rangle + \langle w_\xi, \mathbb{E}\{x_\xi^1\} \rangle\}$$

such that  $x_\xi^1 \in C_1, \quad x_\xi^2 \in C_2(\xi, x_\xi^1)$

# Price of Nonanticipativity

## Here-&-now

$$\min \mathbb{E}\{f(\xi, x^1, x_\xi^2)\}$$

$$x^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x^1), \forall \xi.$$

## Explicit non-anticipativity

$$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2)\}$$

$$x_\xi^1 \in C^1 \subset \mathbb{R}^n,$$

$$x_\xi^2 \in C^2(\xi, x_\xi^1), \forall \xi.$$

$$x_\xi^1 = \mathbb{E}\{x_\xi^1\} \quad \forall \xi$$

$$w_\xi \perp \text{subspace of constant fcns}$$

multipliers

$$\Rightarrow \mathbb{E}\{w_\xi\} = 0$$

$$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2) - \langle w_\xi, x_\xi^1 \rangle + \langle w_\xi, \mathbb{E}\{x_\xi^1\} \rangle\}$$

$$\text{such that } x_\xi^1 \in C_1, \quad x_\xi^2 \in C_2(\xi, x_\xi^1)$$

# Adjusted Here-&-Now

$\min \mathbb{E}\{f(\xi, x^1, x_\xi^2)\}$  such that  $x^1 \in C^1 \subset \mathbb{R}^n$ ,  $x_\xi^2 \in C^2(\xi, x^1)$ ,  $\forall \xi$

$x^1$  must be  $\mathcal{G}$ -measurable,  $\mathcal{G} = \sigma\{\emptyset, \Xi\}$

$x^2$  is  $\mathcal{A}$ -measurable,  $\mathcal{A} \supset \mathcal{G}$ ,

in general, interchange  $\mathbb{E}$  &  $\partial$  is not valid

required:  $\forall \xi, x^1 \in C^1, C^2(\xi, x^1) \neq \emptyset$   $\mathcal{G}$ -measurability of constraints

Now, suppose  $w_\xi$  are the (optimal) non-anticipativity multipliers (prices)

$\min \mathbb{E}\{f(\xi, x_\xi^1, x_\xi^2) - \langle w_\xi, x_\xi^1 \rangle + \langle w_\xi, \mathbb{E}\{x_\xi^1\} \rangle\}$

such that  $x_\xi^1 \in C^1 \subset \mathbb{R}^n$ ,  $x_\xi^2 \in C^2(\xi, x_\xi^1)$ ,  $\forall \xi$

Interchange is now O.K.,  $\mathbb{E}\{\langle w_\xi, \mathbb{E}\{x_\xi^1\} \rangle\} = \langle \mathbb{E}\{w_\xi\}, \mathbb{E}\{x_\xi^1\} \rangle = 0$ , yields

$\forall \xi$ , solve:  $\min f(\xi, x^1, x^2) - \langle w_\xi, x^1 \rangle$  s.t.  $x^1 \in C^1$ ,  $x^2 \in C^2(\xi, x^1)$

a collection of deterministic optimization problems in  $\mathbb{R}^{n_1+n_2}$

# Progressive Hedging Algorithm

0.  $w_\xi^0$  such that  $\mathbb{E}\{w_\xi^0\} = 0$ ,  $v = 0$ . Pick  $\rho > 0$

1. for all  $\xi$ :

$$(x_\xi^{1,v}, x_\xi^{2,v}) \in \arg \min f(\xi; x^1, x^2) - \langle w_\xi^v, x^1 \rangle$$

$$x^1 \in C^1 \subset \mathbb{R}^{n_1}, x^2 \in C^2(\xi, x^1) \subset \mathbb{R}^{n_2}$$

2.  $\bar{x}^{1,v} = \mathbb{E}\{x_\xi^{1,v}\}$ . Stop if  $|x_\xi^{1,v} - \bar{x}^{1,v}| = 0$  (approx.)

otherwise  $w_\xi^{v+1} = w_\xi^v + \rho [x_\xi^{1,v} - \bar{x}^{1,v}]$ , return to 1. with  $v = v + 1$

# Progressive Hedging Algorithm

0.  $w_\xi^0$  such that  $\mathbb{E}\{w_\xi^0\} = 0$ ,  $v = 0$ . Pick  $\rho > 0$

1. for all  $\xi$ :

$$(x_\xi^{1,v}, x_\xi^{2,v}) \in \arg \min f(\xi; x^1, x^2) - \langle w_\xi^v, x^1 \rangle$$

$$x^1 \in C^1 \subset \mathbb{R}^{n_1}, x^2 \in C^2(\xi, x^1) \subset \mathbb{R}^{n_2}$$

2.  $\bar{x}^{1,v} = \mathbb{E}\{x_\xi^{1,v}\}$ . Stop if  $|x_\xi^{1,v} - \bar{x}^{1,v}| = 0$  (approx.)

otherwise  $w_\xi^{v+1} = w_\xi^v + \rho [x_\xi^{1,v} - \bar{x}^{1,v}]$ , return to 1. with  $v = v + 1$

Convergence: add a proximal term

$$f(\xi; x^1, x^2) - \langle w_\xi^v, x^1 \rangle - \frac{\rho}{2} |x^1 - \bar{x}^{1,v}|^2$$

linear rate in  $(x^{1,v}, w^v)$  ... eminently parallelizable

# Nonanticipativity

Recall  $\min Ef(x) = \mathbb{E}\{f(\xi, x(\xi))\}$  such that  $x(\xi) = \mathbb{E}\{x(\xi)\}$   $P$ -a.s.

**Nonanticipativity constraints:**

$\mathcal{N}_a = \{x : \Xi \rightarrow \mathbb{R}^n\} \subset$  linear subspace of constant fcns  
 $\Rightarrow \exists w : \Xi \rightarrow \mathbb{R}$  “multipliers”  $\perp \mathcal{N}_a$  ( $\Rightarrow \mathbb{E}\{w(\xi)\} = 0$ ) such that

$$x^* \in \operatorname{argmin} Ef \implies x^* \in \operatorname{argmin} \{\mathbb{E}\{f(\xi, x(\xi)) + \langle w(\xi), (x(\xi) - \mathbb{E}\{x(\xi)\}) \rangle\}\}$$

$$\implies x^* \in \operatorname{argmin} \{\mathbb{E}\{f(\xi, x(\xi)) + \langle w(\xi), x(\xi) \rangle\}\}$$

$$P\text{-a.s.} \implies x^* \in \operatorname{argmin}_{x \in E} \{f(\xi, x) + \langle w(\xi), x \rangle\}, \quad \xi \in \Xi$$

$w(\cdot)$ : contingencies equilibrium prices,  $\sim$  ‘insurance’ prices

# PH: Implementation issues

implementation: choice of  $\rho$  ... scenario ( $\times$ ), decision (+) dependent  
(heuristic) extension to problems with integer variables

non-convexities: e.g. ground-water remediation with non-linear PDE recourse

asynchronous

partitioning (= different information feeds)

$$\min \mathbb{E}\{f(\xi, x)\}, \quad f(\xi, x) = f_0(x) + l_{C(\xi, x)}(x)$$

$S = \{\Xi_1, \Xi_2, \dots, \Xi_K\}$  a partitioning of  $\Xi$ ,  $p_k = P(\Xi_k)$

$$\mathbb{E}\{f(\xi, x)\} = \sum_n p_n \mathbb{E}\{f(\xi, x) \mid \Xi_n\} \quad (\text{Bundling})$$

defining  $g(k, x) = \mathbb{E}\{f_0(\xi, x) \mid \Xi_n\}$  if  $x \in C_k = \bigcap_{\xi \in \Xi_k} C_\xi$

solve the problem as:  $\min \sum_{n=1}^N p_k g(k, x)$

# Multistage Stochastic Programs

$$\min_{x \in \mathcal{N}^a} \mathbb{E}\{f(\xi, x(\xi))\}, \quad x(\xi) = (x^1(\xi), \dots, x^T(\xi))$$

filtration :  $\mathcal{A}_0 \subset \mathcal{A}_1 \subset \dots \subset \mathcal{A}_T = \mathcal{A}$ ,  $\mathcal{A}_0$  trivial

$x \in \mathcal{N}^a$  if  $x^t$   $\mathcal{A}_{t-1}$ -measurable  $\approx \sigma\text{-field}(\overset{\rightarrow}{\xi}^{v-1})$

(here  $\xi^0$  deterministic,  $x^1(\xi) \equiv x^1$ )

under usual C.Q. (convex case):  $\bar{x} \in \mathcal{X}$  optimal if

$\exists \bar{w} \perp \mathcal{N}^a, \bar{w} \in \mathcal{X}^*$  such that  $\bar{x} \in \arg \min_{x \in \mathcal{X}} Ef(x) - \mathbb{E}\{\langle \bar{w}, x \rangle\}$

$\bar{w} \perp \mathcal{N}^a \Leftrightarrow \mathbb{E}\{\bar{w}(\xi) | \mathcal{A}_{t-1}\} = 0, \forall t = 1, \dots, T$

$\bar{w}$  non-anticipativity prices

at which to buy the right to adjust decision (after observation)  
can be viewed as insurance premiums, ....

# A dual PH-strategy

## single-stage case

minimize  $Ef(x) := \sum_{\xi \in \Xi} p_\xi f(\xi, x)$  over all  $x \in \mathbb{R}^n$

**Strategy:** better estimates of the  $w$ -variables  
“aggregation” of the solutions” to

minimize  $f^a(\xi, x)$  over all  $\xi \in \mathbb{R}^n$  for fixed  $\xi \in \Xi$

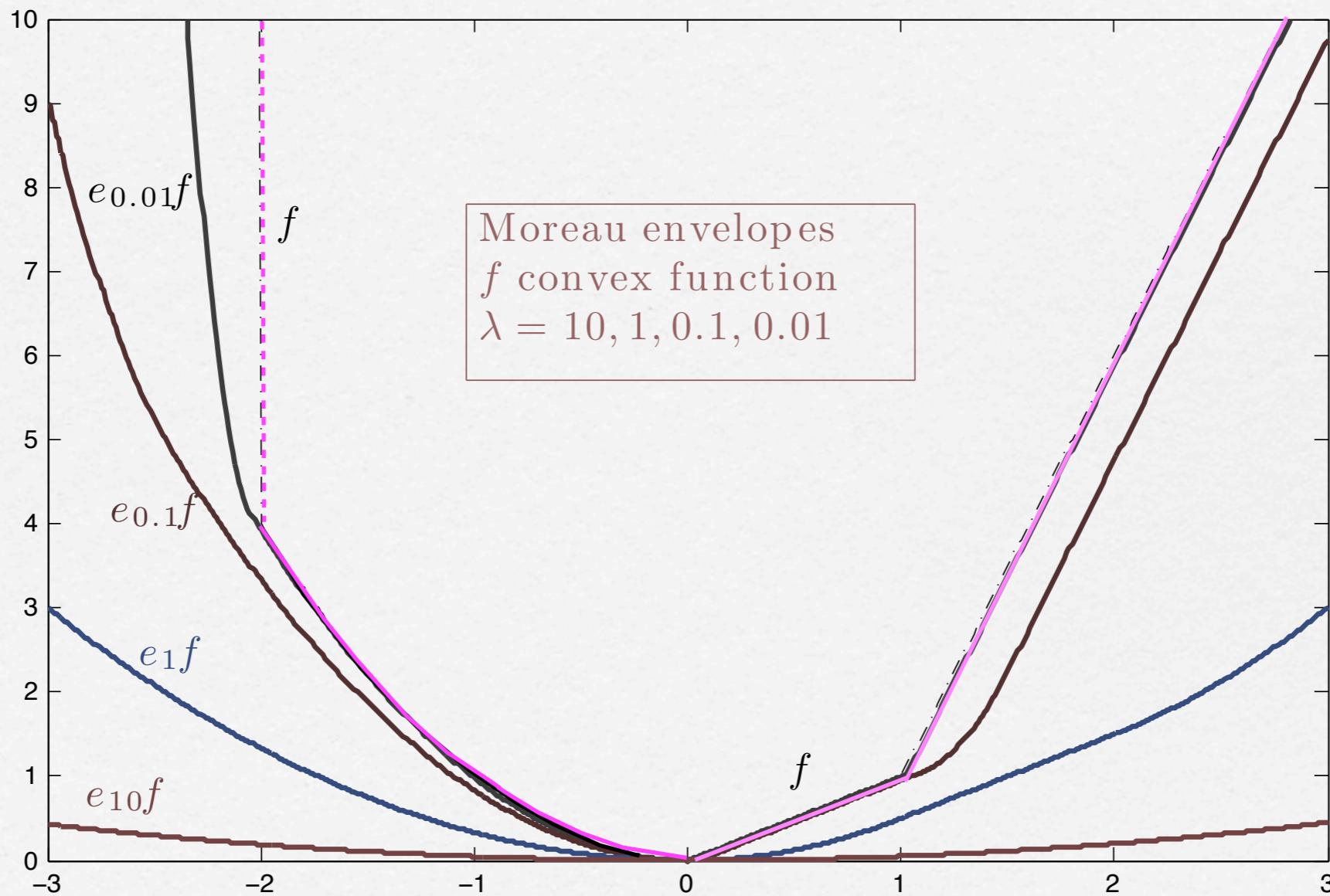
where  $f^a$  approximates  $f$ .

# Moreau envelopes

a.k.a. Moreau-Yosida approximations

$$\text{epi } f \# \text{ epi } g = \inf_u \{f(u) + g(u - x)\} \quad e_\lambda f(x) \text{ with } g = \frac{1}{2\lambda} |\cdot|^2$$

**epi-sums**



# Approximating problem

$$\min F_\lambda(x) := \sum_{\xi \in \Xi} p_\xi f_\lambda(\xi, x), x \in \mathbb{R}^n$$

$$f_\lambda(\xi, x) = \inf_u \left\{ f(\xi, u) + \frac{1}{2\lambda} |u - x|^2 \right\}$$

$F_\lambda \neq (F)_\lambda$  but  $F_\lambda \xrightarrow{p} F, F_\lambda \xrightarrow{e} F$ , finite-valued

**Dual:**  $\max G_\lambda(w) = - \sum_{\xi \in \Xi} p_\xi f_\lambda^*(\xi, w_\xi),$   
such that  $\sum_{\xi \in \Xi} p_\xi w_\xi = 0$

Solution strategy:

$$\min \sum_{k=0}^{\nu-1} \beta_k \alpha_k \text{ such that } \sum_{k=0}^{\nu-1} \beta_k \bar{w}^k = 0$$

$$\sum_{k=0}^{\nu-1} \beta_k = 1, \beta_k \geq 0, k = 0, \dots, \nu - 1$$

for a 'desirable' collection of  $\{\bar{w}^k\}$

Check for optimality, if not, generate  $\bar{w}^\nu$

# Algorithmic procedure

**Step 0.** Initialize by setting  $\nu = 1$  picking  $\{\widehat{w}_\xi^0\}_{\xi \in \Xi}$  in such a way that

$$\bar{w}^0 := \sum p_\xi \widehat{w}_\xi^0 = 0 \text{ & } \alpha_0 = \sum_{\xi \in \Xi} p_\xi [\lambda/2 |\widehat{w}_\xi^0|^2 + \sup_u (\widehat{w}_\xi^0 u - f(\xi, u))]$$

**Step 1.**  $\min_{\beta} \sum_{k=0}^{\nu-1} \beta_k \alpha_k$ , such that  $\sum_{k=0}^{\nu-1} \beta_k \bar{w}^k = 0$ ,

$$\sum_{k=0}^{\nu-1} \beta_k = 1, \quad \beta_k \geq 0, \quad k = 0, \dots, \nu - 1$$

Let  $(z^\nu, \theta^\nu) \in \mathbb{R}^{n+1}$  be the associated multipliers

**Step 2.** For each  $\xi \in \Xi$ , let

$$u_\xi^\nu \in \operatorname{argmin}_u \left\{ f(\xi, u) + \frac{1}{2\lambda} |z^\nu - u|^2 \right\}$$

$$\widehat{w}_\xi^\nu = \lambda^{-1} (z^\nu - u_\xi^\nu), \quad \bar{w}^\nu = \lambda^{-1} \sum_{\xi \in \Xi} p_\xi (z^\nu - u_\xi^\nu)$$

$$\alpha_\xi^\nu = z^\nu \widehat{w}_\xi^\nu - f(\xi, z^\nu - \lambda \widehat{w}_\xi^\nu) - \frac{\lambda}{2} |\widehat{w}_\xi^\nu|^2, \quad \alpha_\nu = \sum_{\xi \in \Xi} p_\xi \alpha_\xi^\nu$$

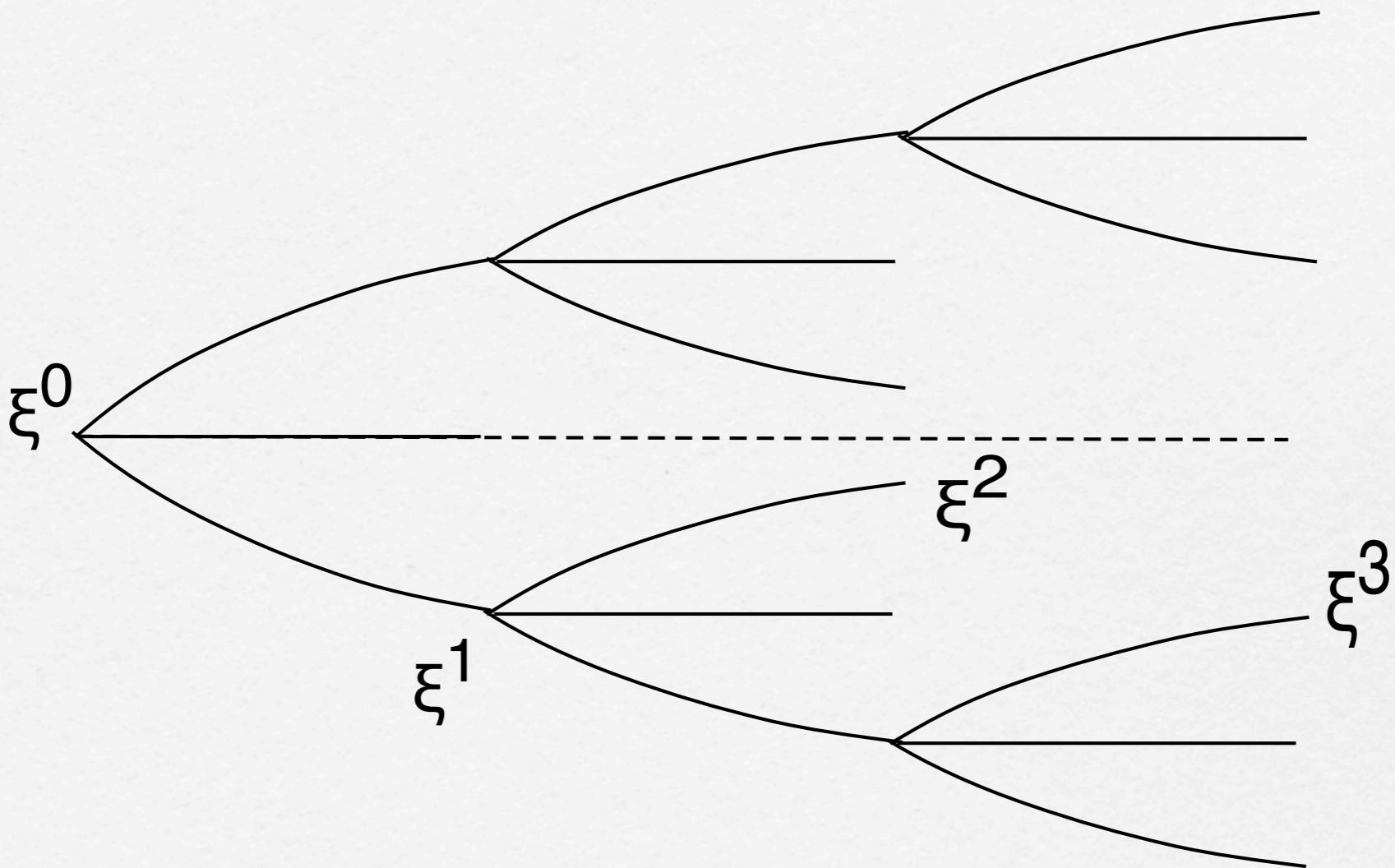
**Step 3.** If  $\alpha_\nu < \bar{w}^\nu z^\nu + \theta^\nu$  return to Step 1 with  $\nu + 1 \rightarrow \nu$

If  $\alpha_\nu \geq \bar{w}^\nu z^\nu + \theta^\nu$ ;  $z^\nu$  is optimal

**Adjust**  $\lambda$  if appropriate; always generates bounds for original problem.

# Bundling

# Discrete Scenario Tree



Algorithm: (1) Nested Sequential SLP  
(2) Progressive Hedging + Bundling

# Bundling Decomposition

$$\min \mathbb{E}\{f(\xi, x)\}, \quad f + \iota_C(\xi), \quad \xi \in \Xi$$

$\iota_C(x) = 0$  when  $x \in C$ ,  $= \infty$  otherwise

$\Xi$  discrete or discretization (not based on best approximation of  $\Xi$ )

$\{\Xi_k, k = 1, \dots, K\}$  a partition of  $\Xi$

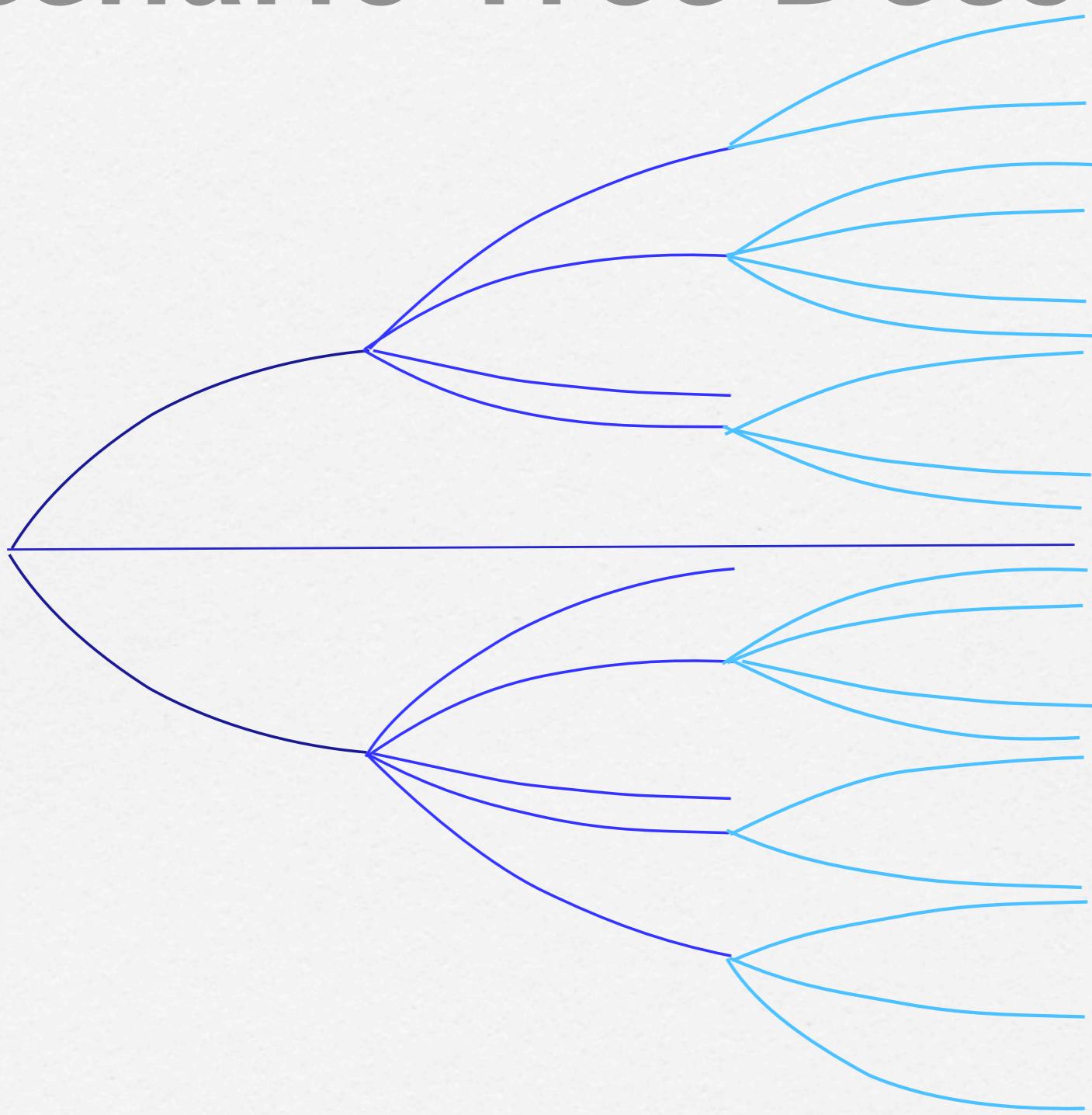
$$p_k = \int_{\Xi_k} P(d\xi), \quad k = 1, \dots, K$$

$$\begin{aligned}\mathbb{E}\{f(\xi, x)\} &= \sum_{k=1}^K p_k \mathbb{E}\{f(\xi, x) \mid \Xi_k\} \\ g(k, x) &= \mathbb{E}\{f(\xi, x) \mid \Xi_k\}\end{aligned}$$

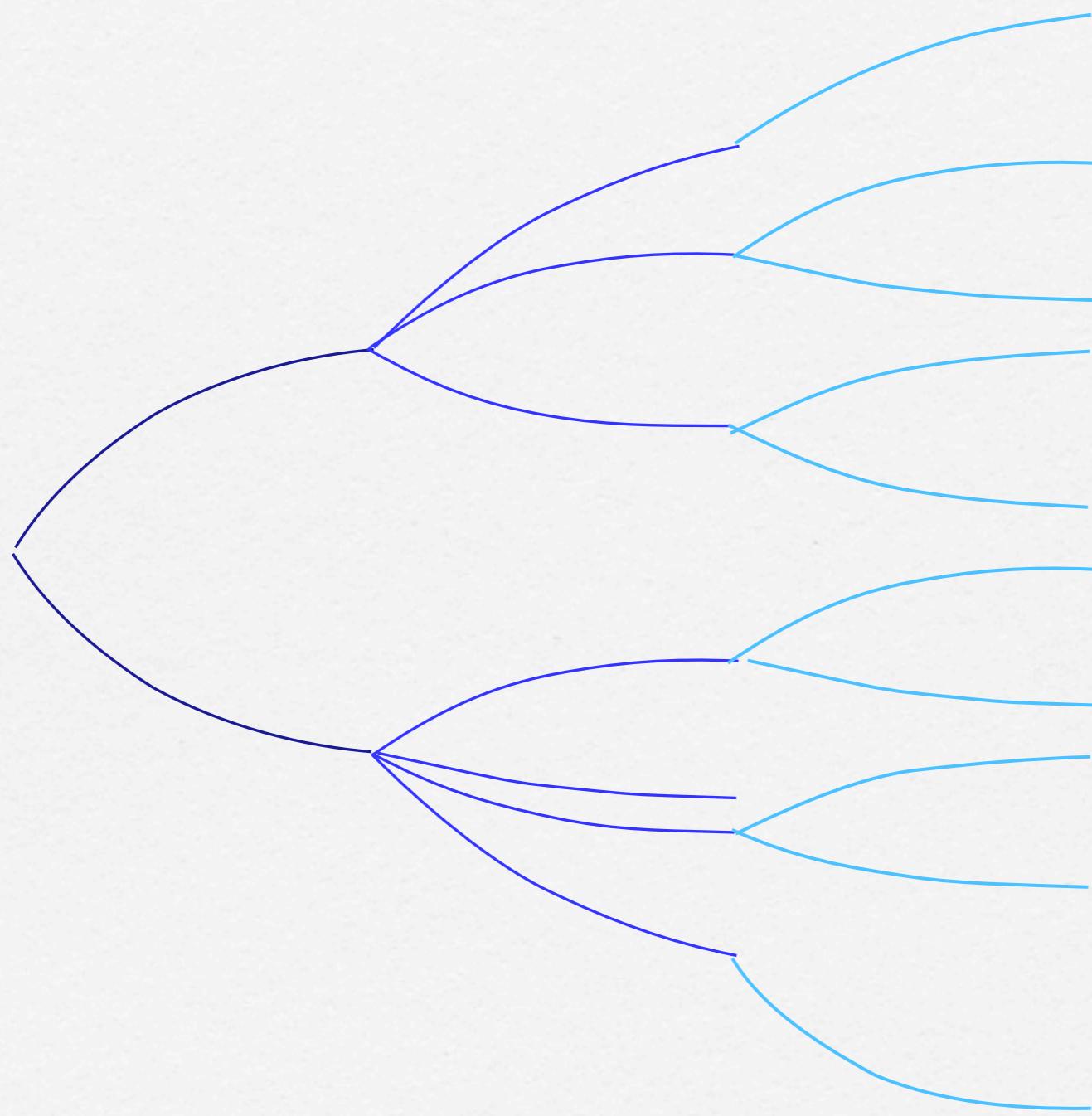
$$\min \sum_{k=1}^K p_k g(k, x)$$

Bayes' Rule

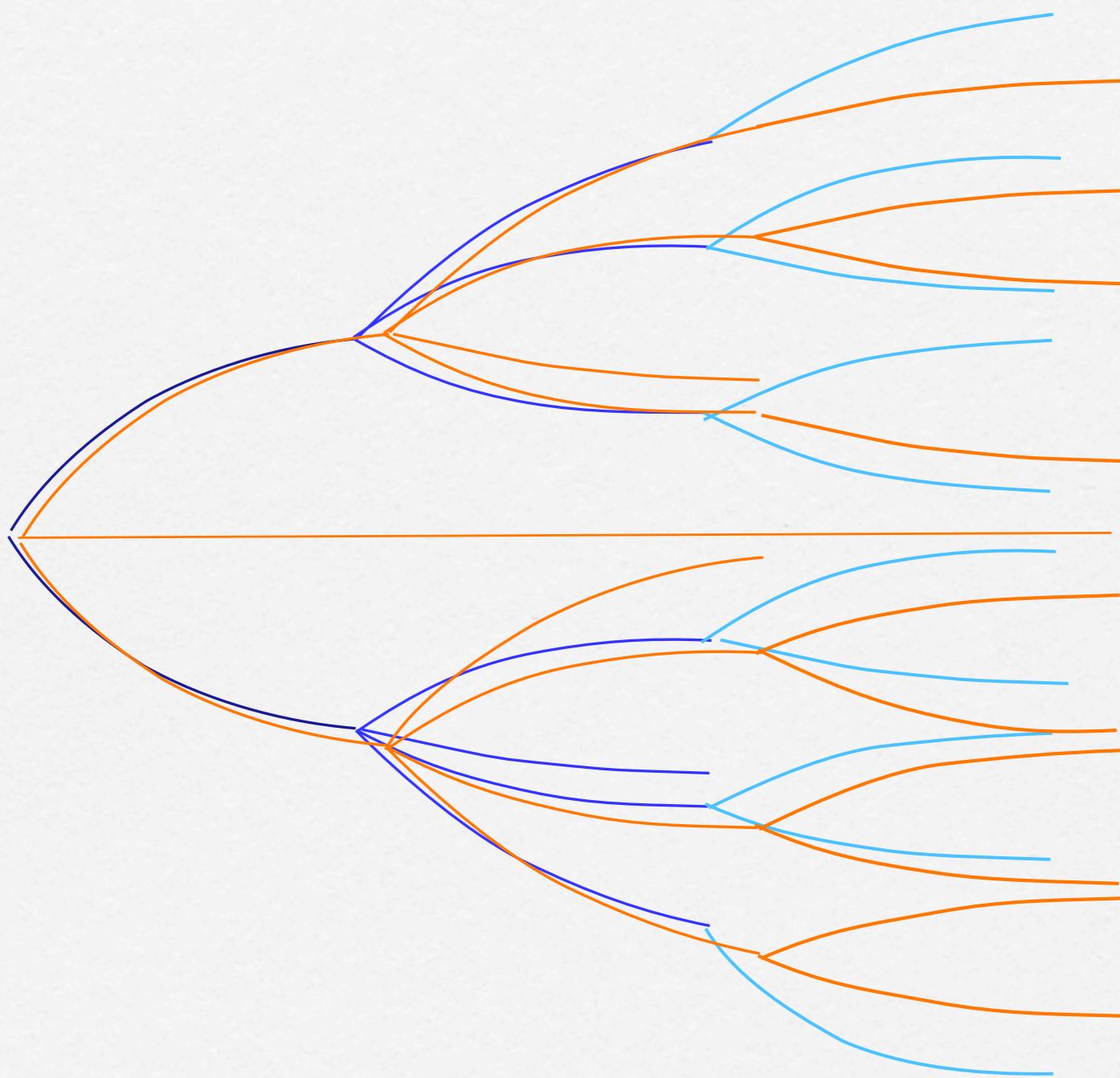
# Scenario Tree Decomposition



# Scenario Tree Decomposition



# Scenario Tree Decomposition



$E_1, E_2$

# ★ PH with Bundling ★

0.  $w_k^0$  such that  $\mathbb{E}\{w_k^0\} = 0$ ,  $v = 0$ . Pick  $\rho > 0$

1. for all  $k : \hat{x}^1, \hat{x}^2 = (x_\xi^2, \xi \in \Xi_k)$

$$(\hat{x}_k^{1,v}, x_k^{2,v}) \in \arg \min g(k; \hat{x}^1, \hat{x}^2) - \langle w_k^v, \hat{x}^1 \rangle$$

$$\hat{x}^1 \in C^1 \subset \mathbb{R}^{n_1}, x^2 \in C^2(\xi, \hat{x}^1) \subset \mathbb{R}^{n_2}$$

2.  $\bar{x}^{1,v} = \sum_{k=1}^K p_k \hat{x}_k^{1,v}$ . Stop if  $|\hat{x}_k^{1,v} - \bar{x}^{1,v}| = 0$  (approx.)

otherwise  $w_k^{v+1} = w_k^v + \rho [\hat{x}_k^{1,v} - \bar{x}^{1,v}]$ , return to 1. with  $v = v + 1$

# ★ PH with Bundling ★

0.  $w_k^0$  such that  $\mathbb{E}\{w_k^0\} = 0$ ,  $v = 0$ . Pick  $\rho > 0$

1. for all  $k : \hat{x}^1$ ,  $\hat{x}^2 = (x_{\xi}^2, \xi \in \Xi_k)$

$$(\hat{x}_k^{1,v}, x_k^{2,v}) \in \arg \min g(k; \hat{x}^1, \hat{x}^2) - \langle w_k^v, \hat{x}^1 \rangle$$

$$\hat{x}^1 \in C^1 \subset \mathbb{R}^{n_1}, x^2 \in C^2(\xi, \hat{x}^1) \subset \mathbb{R}^{n_2}$$

2.  $\bar{x}^{1,v} = \sum_{k=1}^K p_k \hat{x}_k^{1,v}$ . Stop if  $|\hat{x}_k^{1,v} - \bar{x}^{1,v}| = 0$  (approx.)

otherwise  $w_k^{v+1} = w_k^v + \rho [\hat{x}_k^{1,v} - \bar{x}^{1,v}]$ , return to 1. with  $v = v + 1$

Convergence: add a proximal term

$$f(\xi; \hat{x}^1, \hat{x}^2) - \langle w_k^v, \hat{x}^1 \rangle - \frac{\rho}{2} |\hat{x}^1 - \bar{x}^{1,v}|^2$$

linear rate in  $(\hat{x}^{1,v}, w^v)$  ... still eminently parallelizable

# Decentralization?

Aditya Mahajan

