Regaining tractability in some large scale/uncertain engineering optimization problems

Aharon Ben-Tal

Technion – Israel Institute of Technology

THE POWER OF DUALITY IN STRUCTURAL DESIGN

Application Example: Free Material Optimization

♣ FMO is a methodology for design of mechanical structures. In FMO, one seeks how to distribute a given amount of elastic material over a given domain in order to get a structure capable to withstand best of all a given collection of external loads. It is assumed that

- the mechanical properties of the material (its rigidity tensor) may vary, in an arbitrary fashion, from point to point;
- the rigidity of a construction w.r.t. a given external load is measured by the compliance – potential energy capacitated by the construction at the static equilibrium corresponding to the load;

♠ The goal is, given the weight of the construction, to minimize its largest, over a given set of loading scenarios, compliance.

♣ Usually it is technically impossible or too expensive to implement an FMO design "as it is". The role of FMO is in providing a good guess for the structure of the would-be construction. After the structure is guessed, the construction is designed from traditional materials via standard engineering techniques.

♣ **With Finite Element discretization, the Multi-Load FMO problem is**

$$\min_t \left\{ \max_{\ell=1,\dots,K} f_\ell^T S^{-1}(t) f_\ell : t_i \succeq 0, \sum_i \mathrm{Tr}(t_i) \le 1 \right\} \qquad \textbf{(FMO)}$$

where

- t_i, $i = 1, ..., N$, **are symmetric** 3×3 **(in 2D) or** 6×6 **(in 3D) variable matrices (rigidity tensors of the material in Finite Element cells),**
- f_ℓ, $\ell = 1, ..., K$, **are** M**-dimensional data vectors representing loading scenarios,**
- $S(t) = \sum_{i,s} b_{is}^T t_i b_{is}$ **is the** $M \times M$ **<u>stiffness matrix</u> of the construction.**

♣ **In a realistic 2D FMO problem,**

- **the number** N **of Finite Element cells is tens of thousands**

$\Rightarrow$ **design dimension of (FMO) is of order of 50,000 – 200,000**

- **the size** M **of the stiffness matrix is** $\approx 2N$

$\Rightarrow$ **it is a nontrivial problem just to compute the objective!**

$$\text{(FMO)} \qquad \min_{t_i \in S_{\dim}} \max_{\ell=1\ldots k} \left\{ \frac{1}{2} f_\ell^T S^{-1}(t) f_\ell \;\middle|\; \begin{array}{l} t_i \succcurlyeq 0 \\ \Sigma Tr(t_i) \preccurlyeq w \end{array} \right\}$$

$$S(t) = \sum_{i=1}^{m} \sum_{j=1}^{d} b_{ij} t_i b_{ij}^T$$

$$\Leftrightarrow \qquad \min_{t,\tau} \left\{ \tau \;\middle|\; \frac{1}{2} f_\ell^T S^{-1}(t) f_\ell \leq \tau, \;\; \begin{array}{l} t_i \succcurlyeq 0 \\ \Sigma Tr(t_i) \preccurlyeq w \end{array} \right\}$$

$\Leftrightarrow$ Schur complement

$$\min \tau$$

$$\begin{array}{c} (m+1)\times(m\times 1) \\ \text{matrix} \end{array} \quad \begin{pmatrix} 2\tau & -f_\ell^T \\ -f_\ell & \Sigma\Sigma b_{ij} t_i b_{ij}^T \end{pmatrix} \succcurlyeq 0 \quad \ell = 1,\ldots K$$

$$\text{spectrahedron} \qquad \Sigma Tr(t_i) \leq w, \quad t_i \succcurlyeq 0$$

$$\tau \in \mathbb{R}.$$

- The dual problem is

$$-2\sum_{\ell=1}^{k} f_\ell^T v_\ell + w\gamma \to \min$$

$$\left(\begin{array}{ccc|c|ccc|c}
\alpha_1 & & & & & & & v_1^T b_{i1} \\
& \ddots & & & & & & \cdots \\
& & \alpha_1 & & & & & v_1^T b_{iS} \\
\hline
& & & \ddots & & & & \cdots \\
\hline
& & & & \alpha_k & & & v_k^T b_{i1} \\
& & & & & \ddots & & \cdots \\
& & & & & & \alpha_k & v_k^T b_{iS} \\
b_{i1}^T v_1 & \cdots & b_{iS}^T v_1 & \cdots & b_{i1}^T v_k & \cdots & b_{iS}^T v_k & \gamma I_3
\end{array}\right) \succeq 0,\ \ i = 1, \ldots, n;$$

$$2\sum_{\ell=1}^{k} \alpha_\ell = 1\,.$$

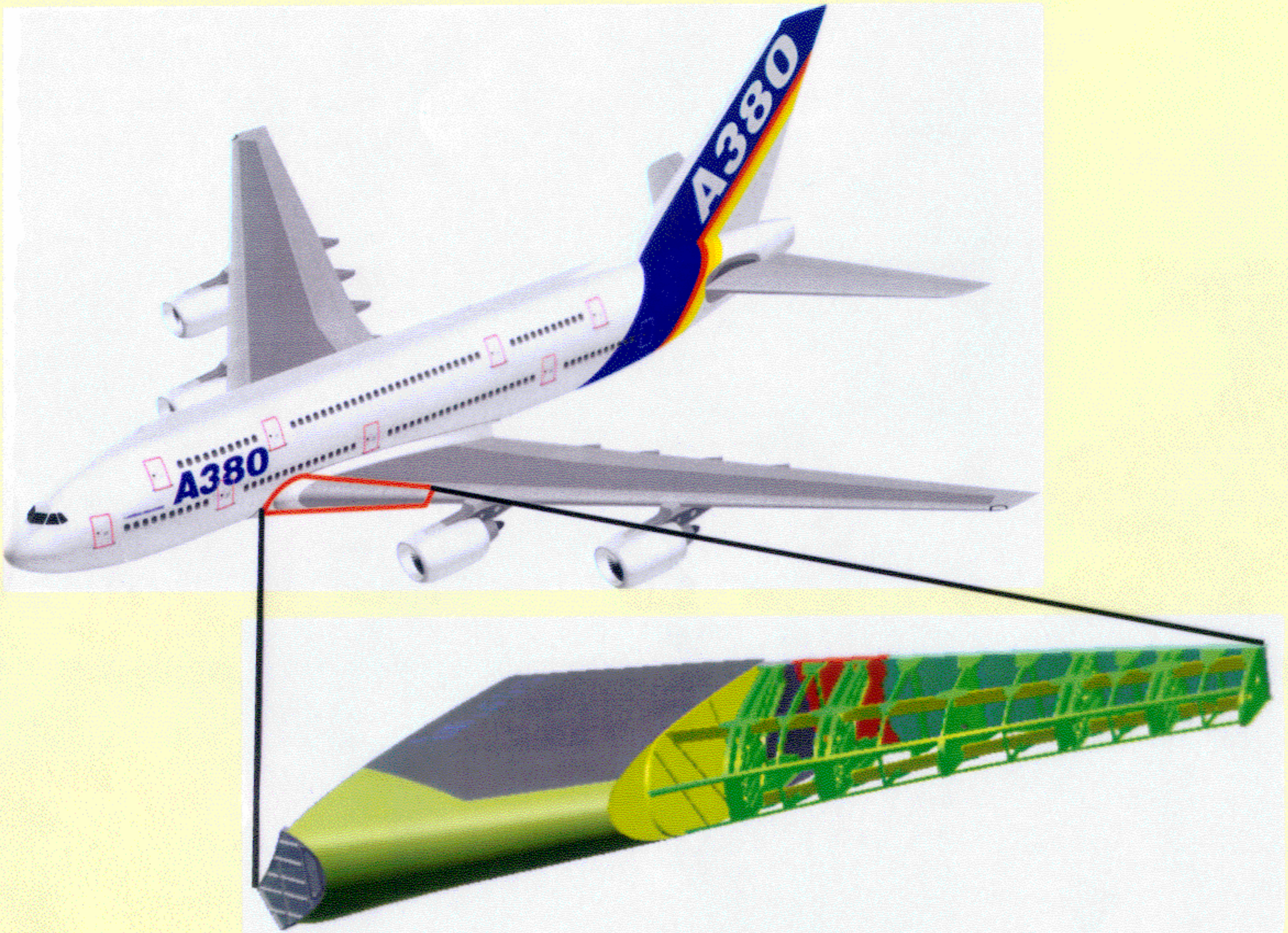
$$[\alpha_\ell,\ \gamma \in \mathbb{R},\ v_\ell \in \mathbb{R}^m]$$

- E.g., for planar shape with 14×14 cells and 3 loads:

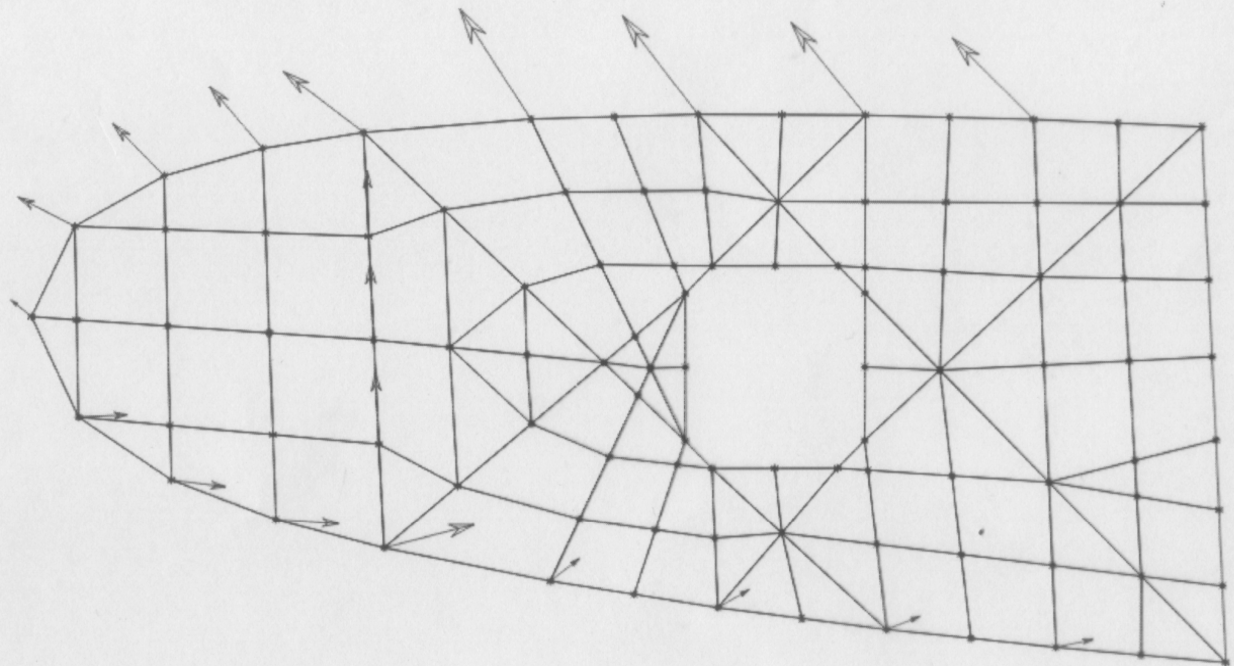
Setting	Design dimension	Effort of analyzing LMI's at a point, a.o.
(Pr)	1,177	37,309,230
$(D1)$	1,264	71,608

Optimization in Flight...

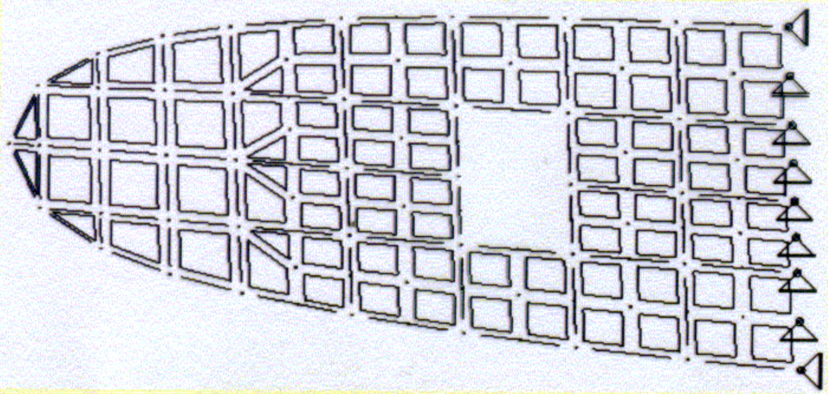
Design of stiffeners: MOPED & MBB-LAGRANGE

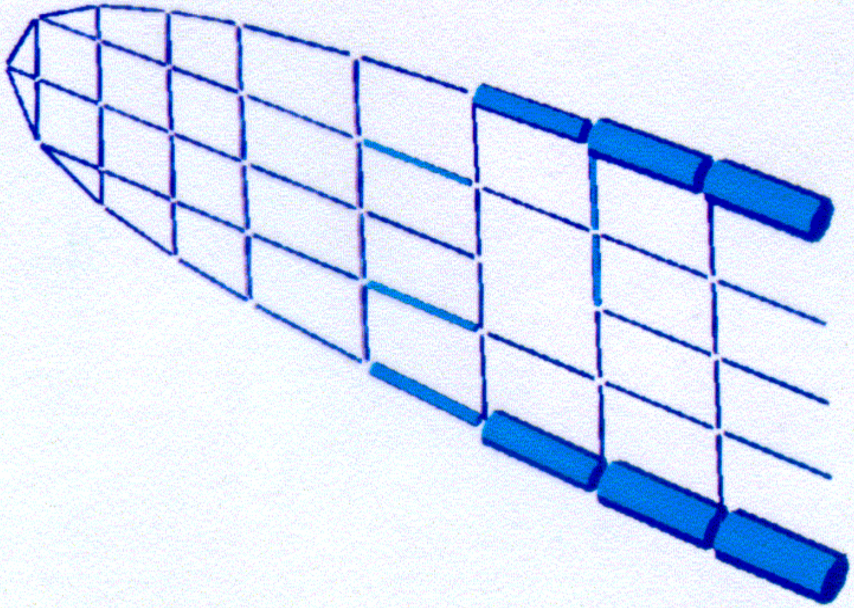


Design of stiffeners: MOPED & MBB-LAGRANGE

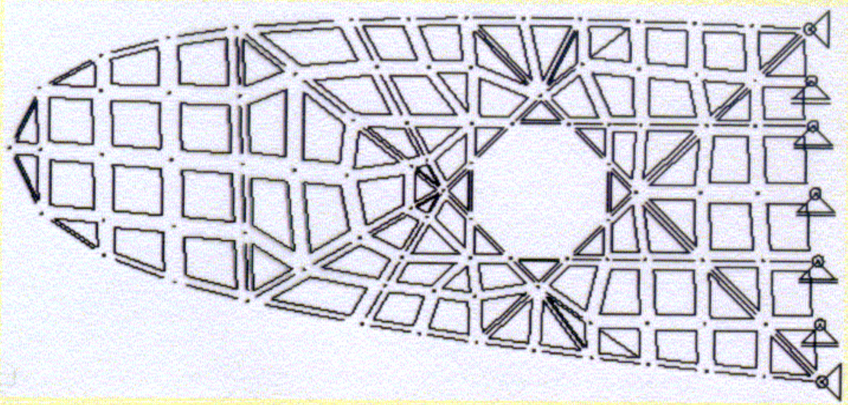


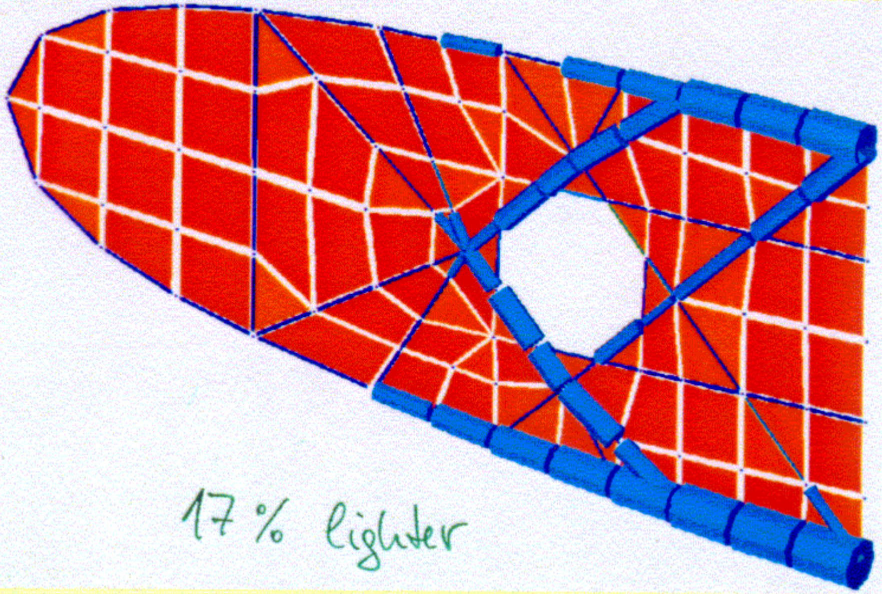
Reference design

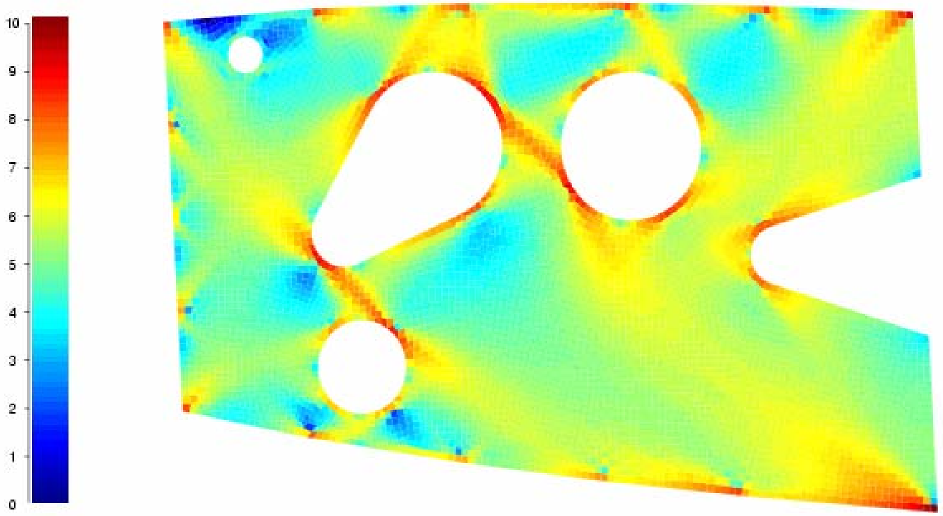




FMO based design







Wing element for Aerobus A380
FMO design by NERML
80 iterations

Implementation, Erlangen University and European Aero Defence and Space Co.

$n = 39,780,\ N = 6,630,\ M = 13,824$

Free Material Optimization: element of aircraft wing
FMO allowed for 17% reduction in element's weight

Based on computational results for maximum stiffness and quite a bit of engineering interpretation a new type of structure was devised for the ribs which gave a weight benefit against traditional and competitive honeycomb/ composite designs (up to 40% !)

A total weight saving of more than 500 kg per wing was obtained by optimizing the ribs in the area shown. These are now — since April 27, 2005 - the first topology optimized parts in flight.

Recovery of signals from noisy outputs

The Estimation Problem

$$y = Hx + w$$

Given y, find an estimator $\hat{x}$, which is as "close" as possible to x.

w random vector

$E(w) = 0, \;\; \mathrm{cov}(w) = C$ positive definite

CLASSICAL METHODS are based on minimizing data error $\|y - Hx\|$.

CLASSICAL APPROACH (Gauss,...)

Closeness measured by (standardized) data error

$$\|C^{-1/2}(y - H\hat{x})\|_2$$

Least Squares Estimator

$$\hat{x}_{LS} = \arg\min_x \|C^{-1/2}(y - Hx)\|_2$$

convex optimization

SOLUTION (H full column rank)

$$\boxed{\hat{x}_{LS} = (H^T C^{-1} H)^{-1} H^T C^{-1} y}$$

a *linear estimator*

$$\hat{x} = Gy$$

CLASSICAL MODIFICATION (Tikhonov,...)

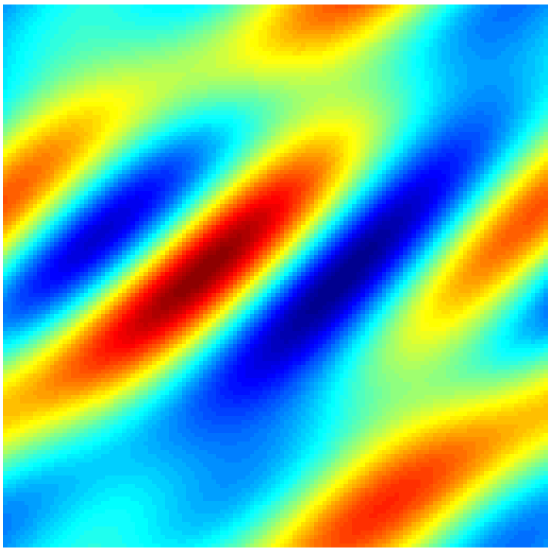
$$\boxed{\hat{x}_T = \arg\min_x \left\{\|C^{-1/2}(y - Hx)\|^2 + \lambda\|x\|^2\right\}}$$

still convex optimization

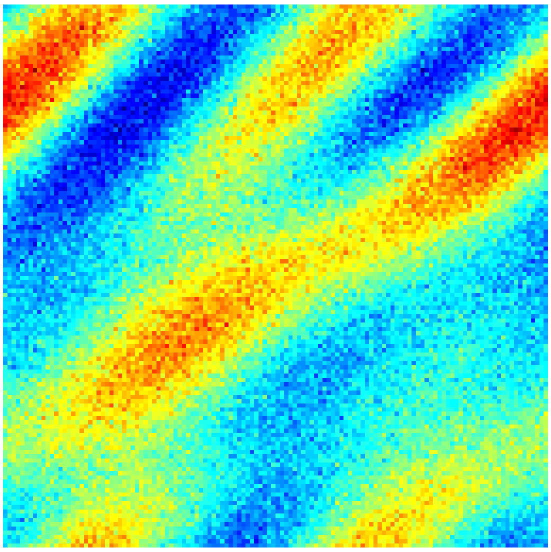
SOLUTION

$$\hat{x}_T = (H^T C^{-1} H + \lambda I)^{-1} H^T C^{-1} y$$

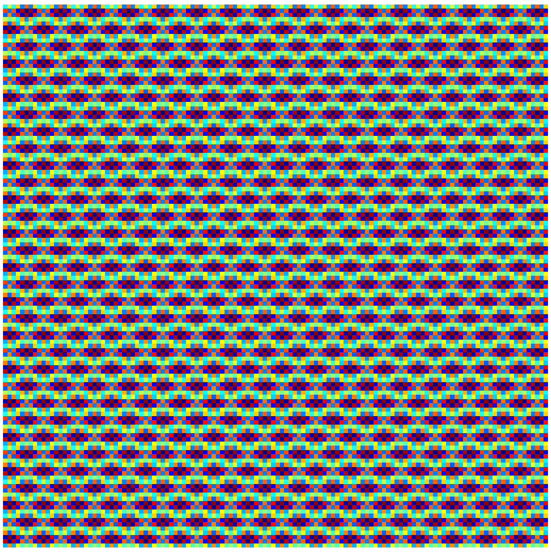
also a linear estimator.



True signal



Observations



LS

MSE estimator

$$\min_{\hat{x}} \; E\|x - \hat{x}\|^2$$

With a *linear estimator* $\hat{x} = Gy$ problem becomes

$$\min_{G} \left\{ x^T (I - GH)^T (I - GH) x + Tr(GCG) \right\}$$

bias variance

but x unknown!

"Solution": minimal variance unbiased estimator $GH = I$

Solution: Same as $\hat{x}_{L_S} \cdots$

Our approach: minmax MSE linear estimator: $\hat{x} = Gy$, where:

$$\min_{G} \max_{\|x\|_T \le L} \left\{ x^T (I - GH)^T (I - GH) x + Tr(GCG) \right\}$$

Employing conic optimization theory we proved:

Theorem I: Original MinMax MSE problem (1) is equivalent to the SDP problem:

$$\min_{s.t.} L^2\lambda + t$$

$$\begin{pmatrix} \lambda I & T^{-1/2}(I - GH)^T \\ (I - GH)T^{-1/2} & I \end{pmatrix} \succeq 0$$

$$\begin{pmatrix} t & g^T \\ g & I \end{pmatrix} \succeq 0$$

Variables G, λ, t

$g = \text{vec}\left(GC^{1/2}\right)$

Theorem II: For the special case $T = I$, SDP can be solved explicitly. The optimal MMX MSE estimator is

$$\hat{x}_{\text{mmx}} = \alpha \underbrace{(H^T C^{-1} H)^{-1} H^T C^{-1} y}_{\hat{x}_{LS}}$$

where $$\alpha = \frac{L^2}{L^2 + Tr((H^T C^{-1} H)^{-1})}$$

Proof Structure

(I) Establish the structure of the optimal solution

$$\boxed{G = VDV^T(H^TC^{-1}H)^{-1}H^TC^{-1}}$$

where V is the orthogonal matrix diagonalizing $H^TC^{-1}H$, i.e.,

$$H^TC^{-1}H = V\Sigma V^*$$

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$$

This is obtained by optimality condition. Using this, we end up with an equivalent problem in variable (matrix) D (Problem B below).

(II) Show that $\exists$ an optimal matrix D which is diagonal.

(III) Find the diagonal elements of D.

(B)

$$\boxed{\begin{array}{c}\min_{D,\lambda} \; L^2\lambda + Tr(D^T D\Sigma^{-1}) \\ (I-D)^T(I-D) \preccurlyeq \lambda I\end{array}}$$

Second part of the proof ("optimal D can be chosen diagonal").

Let $\mathcal{J}_n$ be the set of 2^n matrices which are $n \times n$, diagonal, with the entries in the diagonal being $+1$ or -1.

Claim If D^* is an optimal solution of (B), then so is

$$JD^*J, \quad \forall\, J \in \mathcal{J}_n$$

Proof

$$\begin{array}{rcl} Tr[(JDJ)^T(JDJ)\Sigma^{-1} & = & Tr(D^T D\Sigma^{-1})] \\ (I-JDJ)^T(I-JDJ) \preceq \lambda I & \Leftrightarrow & (I-D)^T(I-D) \preccurlyeq \lambda I \end{array}$$

Conclusion Since (B) is a convex problem $\Rightarrow$ its optimal solution set is convex, so if D^* is an optimal solution, so is

$$\frac{1}{2^n} \sum_{J\in\mathcal{J}_n} (JD^*J)$$

$$D = \begin{bmatrix} a, & b \\ c, & d \end{bmatrix}$$

$$\mathcal{J}_2 = \Bigg\{ \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{J_1}, \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}}_{J_2}, \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}}_{J_3} \underbrace{\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}}_{J_4}$$

$$J_1 D J_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad J_2 D J_2 = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix}$$

$$J_3 D J_3 = \begin{pmatrix} a & -b \\ -c & d \end{pmatrix} \qquad J_4 D J_4 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\frac{1}{4}\sum_{i=1}^{4} J_i D J_i = \frac{1}{4}\begin{bmatrix} 4a & 0 \\ 0 & 4d \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix}$$

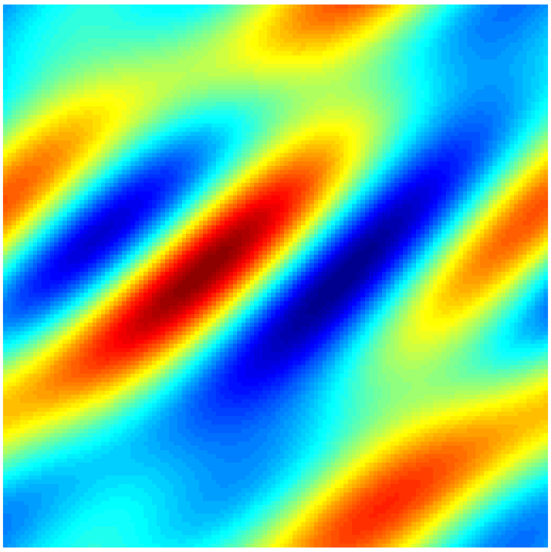
General result

$$\frac{1}{2^n} \sum_{J \in \mathcal{J}_n} JDJ = \text{diag } D$$

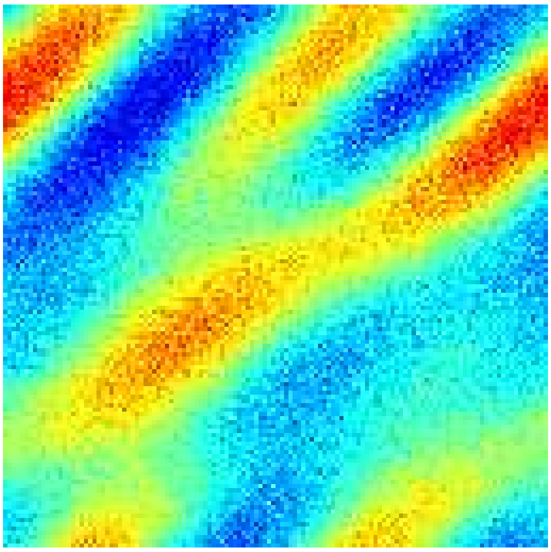
Part 3 of the proof with $D = \text{diag}(d_1, \ldots, d_n)$ problem (A) $\Leftrightarrow$ (B) reduces to

$$\boxed{\begin{array}{ll} \min\limits_{d_i, \lambda} & L^2\lambda + \sum (d_i^2/\sigma_i) \\ \text{s.t.} & (1 - d_i)^2 \le \lambda\,, \quad \forall\ i \end{array}}$$

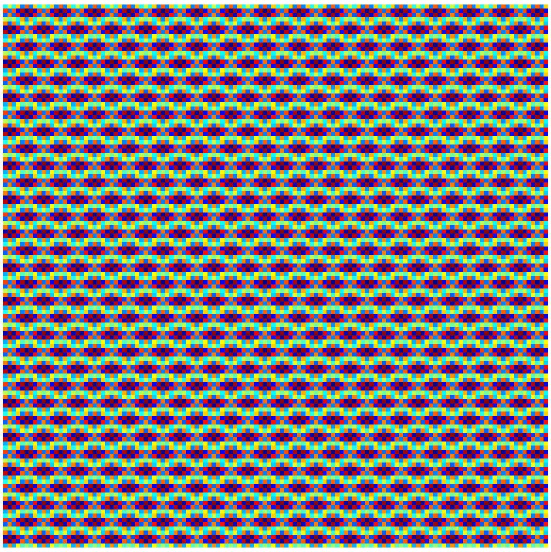
This problem can be solved analytically, which gives the final result claimed in Theorem II.



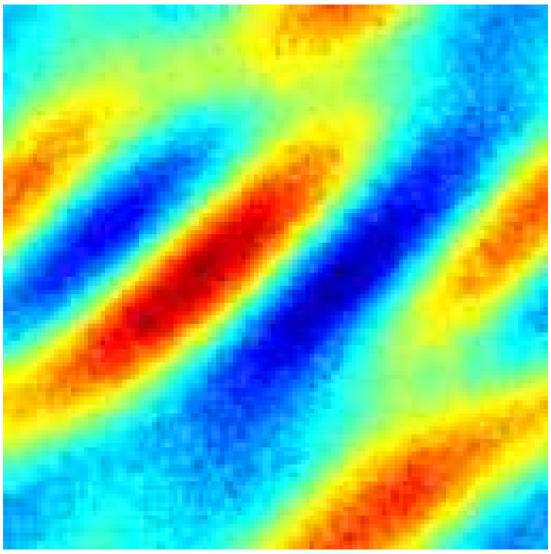
True signal



Observations



LS



Minmax Use

Dynamic Robust Optimization

$$\textbf{open loop dynamics:} \quad \begin{cases} x_{t+1} &= A_t x_t + B_t u_t + R_t d_t \\ y_t &= C_t x_t \\ x_0 &= z \end{cases}$$

$$\textbf{control law:} \quad u_t = \xi_{t0} + \sum_{\tau=0}^{t} \Xi_{t\tau} y_\tau$$

$$\Downarrow$$

$$w := (u_0, ..., u_T, x_0, ..., x_{T+1}) = W(\xi; d, z)$$

$$\Downarrow$$

$$\min_{\xi} \{ f(W(\xi; d, z)) : D_i W(\xi; d, z) - b_i \in \mathcal{Q}_i,\ i = 1, ..., m \} \qquad \textbf{(U)}$$

<u>Note:</u> Due to presence of uncertain input trajectory d and possible uncertainty in the initial state, (U) is an uncertain problem.

<u>Difficulty:</u> While linearity of the dynamics and the control law make $W(\xi; d, z)$ linear in (d, z), the dependence of $W(\cdot, \cdot)$ on the parameters $\xi = \{\xi_{t0}, \Xi_{t\tau}\}_{0 \le \tau \le t \le T}$ of the control law is highly nonlinear

$\Rightarrow$ (U) is *not* a bi-affine problem, which makes inapplicable the theory we have developed. In fact, (U) seems to be intractable already when there is no uncertainty in d, z!

Dynamic Control Problems

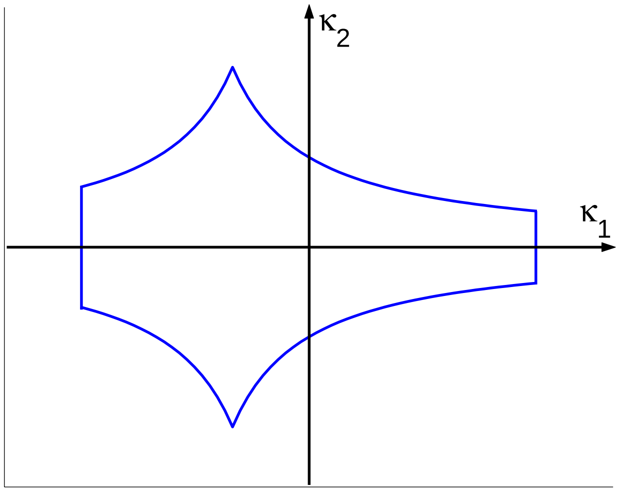
- Example:

$$\mathcal{X}_{t+1} = \mathcal{X}_t + u_t + d_t$$
$$\mathcal{Y}_t = \mathcal{X}_t$$
$$\mathcal{X}_0 = 0$$
$$u_t = \kappa_t \mathcal{Y}_t$$

- Control wanted: when $|d_t| \le 1$ then $|u_t| \le 3$ for t=0,1,2
- Since: $u_2 = \kappa_2 \mathcal{X}_2 = \kappa_2(\mathcal{X}_1 + \kappa_1 d_0 + d_1) = \kappa_2\left[(1+\kappa_1)d_0 + d_1\right]$

  is not bi-affine

  the control coefficients have a highly non convex domain.



<u>Remedy:</u> **suitable re-parameterization of affine control laws.**

♣ **Consider a closed loop system along with its** *model*:

closed loop system:	model:
$x_{t+1} = A_t x_t + B_t u_t + R_t d_t$	$\widehat{x}_{t+1} = A_t \widehat{x}_t + B_t u_t$
$y_t = C_t x_t$	$\widehat{y}_t = C_t \widehat{x}_t$
$x_0 = z$	$\widehat{x}_0 = 0$
$u_t = U_t(y_0, ..., y_t)$	

♠ <u>**Observation:**</u> **We can run the model in an on-line fashion, so that at time** t, **before the decision on** u_t **should be made, we have in our disposal** *purified outputs*

$$v_t = y_t - \widehat{y}_t.$$

♠ <u>**Fact I [Equivalence]:**</u> **Every transformation** $(d, z) \mapsto w$ **which can be obtained from an affine control law based on outputs:**

$$u_t = \xi_{t0} + \sum_{\tau=0}^{t} \Xi_{t\tau} y_\tau \qquad (*)$$

can be obtained from an affine control law based on purified outputs:

$$u_t = \eta_{t0} + \sum_{\tau=0}^{t} H_{t\tau} v_\tau \qquad (**)$$

and vice versa.

system:	**model:**
$x_{t+1} = A_t x_t + B_t u_t + R_t d_t$	$\widehat{x}_{t+1} = A_t \widehat{x}_t + B_t u_t$
$y_t = C_t x_t$	$\widehat{y}_t = C_t \widehat{x}_t$
$x_0 = z$	$\widehat{x}_0 = 0$

(S)

control law:

$$\begin{aligned} v_t &= y_t - \widehat{y}_t \\ u_t &= \eta_{t0} + \sum_{\tau=0}^{t} H_{t\tau} v_\tau \end{aligned} \qquad (**)$$

♠ **Fact II [bi-affinity]: The state-control trajectory $w = W(\eta; d, z)$ of (S) is affine in (d, z) when the parameters $\eta = \{\eta_{t0}, H_{t\tau}\}_{0 \le \tau \le t \le T}$ of the control law $(**)$ are fixed,** *and is affine in η when (d, z) is fixed.*

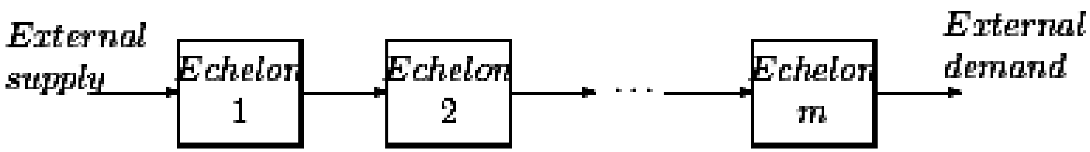
♠ **Corollary:** *With parameterization $(**)$ of affine control laws, the problem*

> *Find an affine control law $(*)$ which ensures that the resulting state-control trajectory w satisfies the system of convex inclusions*
>
> $$D_i w - b_i \in \mathcal{Q}_i,\ i = 1, ..., m$$
>
> *and minimizes, under this restriction, a given linear objective $f(w)$.*

becomes an uncertain bi-affine optimization problem and as such can be processed via the CRC approach.

In particular, in the case when $\mathcal{Q}_i$ are one-dimensional, the CRC of the problem is computationally tractable, provided that the normal range $\mathcal{U}$ of (d, z) and the associated cone $\mathcal{L}$ are so. If $\mathcal{U}$, $\mathcal{L}$ and the norms used to measure distances are polyhedral, CRC is just an explicit LP program.

The Supply chain Problem



- x_t^j=amount echelon j orders from j-1 at the beginning of period t
- Y_t^j =inventory level in echelon *j* at the end of period *t*
- z^j =initial inventory level at echelon *j*
- d_t =external demand at period *t*

- *I(j)*= Information delay, *M(j)*= manufacturing delay, *L(j)*= Lead time
- $T^L(j) = I(j) + M(j-1) + L(j)$ the delay between the time an order is placed and received in echelon *j*
- $T^M(j) = I(j+1) + M(j)$ the delay between the time an order is placed and shipped from echelon j

The Supply chain Problem

- Main objective : minimizing cost
- Sub objective:
  stabilizing the system
- Problem Characteristics:
  - Finite horizon
  - Multi echelon
  - Delays
  - Backlogging
  - Demand must be satisfied
    and is uncertain
- Eliminating the equalities
  recursively yields a LP
  with only inequalities

$$
\begin{aligned}
&\min_{x,y} \sum_{j,t} \left[c_t^j x_t^j + w_t^j \right] \\
&s.t. \\
&\quad y_t^j = y_{t-1}^j + x_{t-T^L(j)}^j - x_{t-T^M(j)}^{j+1} \qquad \forall j \in \{1,..,m-1\} \\
&\quad y_t^m = y_{t-1}^m + x_{t-T^L(j)}^m - d_{t-T^M(m)} \\
&\left.\begin{aligned}
&w_t^j \ge h_t^j y_t^j \\
&w_t^j \ge -p_t^j y_t^j \\
&y_t^j \ge \underline{a}^j \\
&y_t^j \le \overline{a}^j \\
&x_t^j \le b^j \\
&x_t^j \ge 0 \\
&w_t^j \ge 0 \\
&y_0^j = z^j
\end{aligned}\right\} \forall j \in \{1,..,m\}
\end{aligned}
$$

The Bullwhip effect

- This problem has a well known phenomenon associated with it called “Bullwhip effect”
- The Bullwhip effect is described as “amplification of oscillation from down stream demands to upstream echelons”
- Such amplification can occur both in orders and inventory levels.
- Large variations in these measures are disruptive to the system and generates high cost.
- One of the aims of good control is to reduce the Bullwhip effect.

LV control

- Let's assume we take the control suggested by Love [Love,1979] using target inventory:

$$x_t^j = x_{t-1}^{j+1} + \tfrac{1}{2}\left(\Gamma^j - y_{t-1}^j\right)$$

- Resulting in the LP problem :

$$
\begin{aligned}
&\min_{x,\tilde{x},y,w,\pi} \sum_{j,t} \left[c_t^j x_t^j + w_t^j\right] \\
&s.t. \\
&\quad y_t^j = y_{t-1}^j + x_{t-T^L(j)}^j - x_{t-T^M(j)}^{j+1} \qquad \forall j \in \{1,..,m-1\} \\
&\quad y_t^m = y_{t-1}^m + x_{t-T^L(j)}^m - d_{t-T^M(m)} \\
&\left.\begin{aligned}
&w_t^j \ge h_t^j y_t^j \\
&w_t^j \ge -p_t^j y_t^j \\
&\tilde{x}_t^j = x_{t-1}^{j+1} + \tfrac{1}{2}\left(\Gamma^j - y_{t-1}^j\right) \\
&\tilde{x}_t^j \le \mathcal{M}\pi_t^j \\
&x_t^j \ge 0 \\
&x_t^j \ge \tilde{x}_t^j \\
&x_t^j \le \mathcal{M}\pi_t^j \\
&x_t^j \le \mathcal{M}(1-\pi_t^j) + \tilde{x}_t^j \\
&w_t^j \ge 0 \\
&y_0^j = z^j \\
&\pi_t^j \in \{0,1\}
\end{aligned}\right\} \forall j \in \{1,..,m\}
\end{aligned}
$$

- We can further improve this method by making the reference inventory a decision variable rather than a constant. – **ILV control**

Example

- [Love, 1979], Oscillating demand:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
d_t	6	6	6	6	6	6	6	6	7	8	9	10	9	8	7	6	5	4	5	6

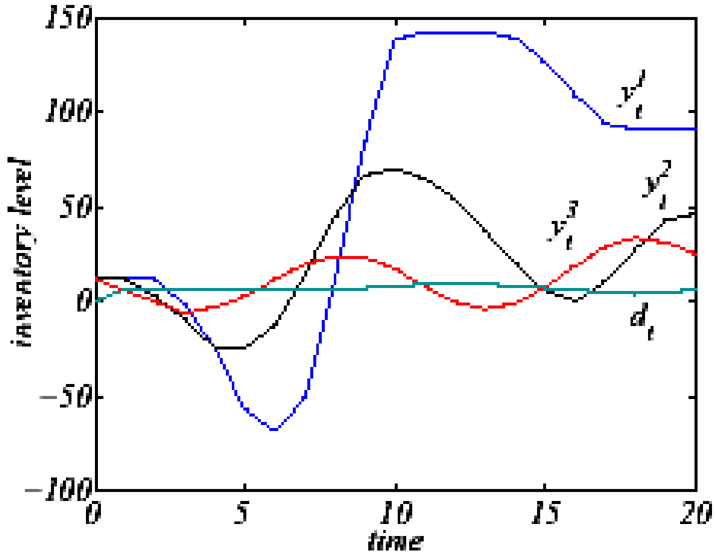
- Horizon: n=20
- Echelons: m=3
- Cost: c=2, p=3,h=1
- Initial inventory: z=12
- Lead time: L=2
- No other delays

The importance of good control

- The Bullwhip effect –

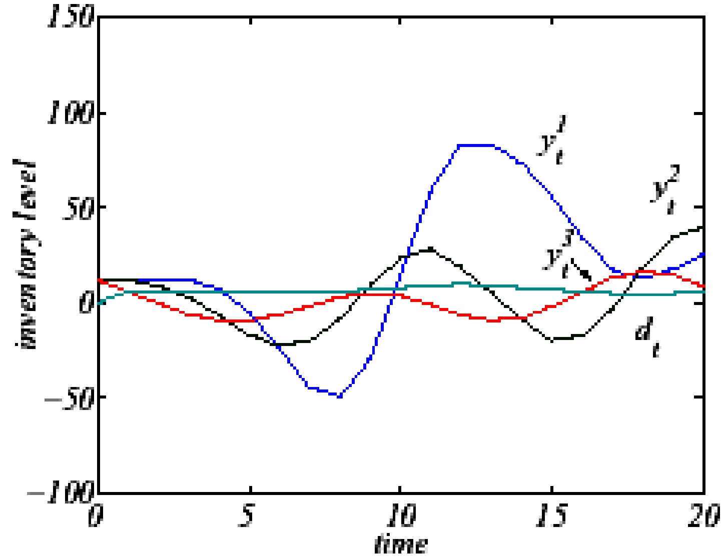
LV

OFV: 4795



ILV

OFV: 3140



- These are deterministic methods which do not work well with varying demand → a more robust method is needed

Purified output-based AARC control

The supply chain problem as a control problem

- The dynamics of the supply chain problem is given by:

$$y_t^j = y_{t-1}^j + x_{t-T^L(j)}^j - x_{t-T^M(j)}^{j+1} \quad \forall j \in \{1,..,m-1\}$$
$$y_t^m = y_{t-1}^m + x_{t-T^L(j)}^m - d_{t-T^M(m)}$$

with initial state: $y_0^j = z^j$

whose form matches the classical dynamic control problem

The purified outputs corresponding to the dynamic system (1) – (3) are here

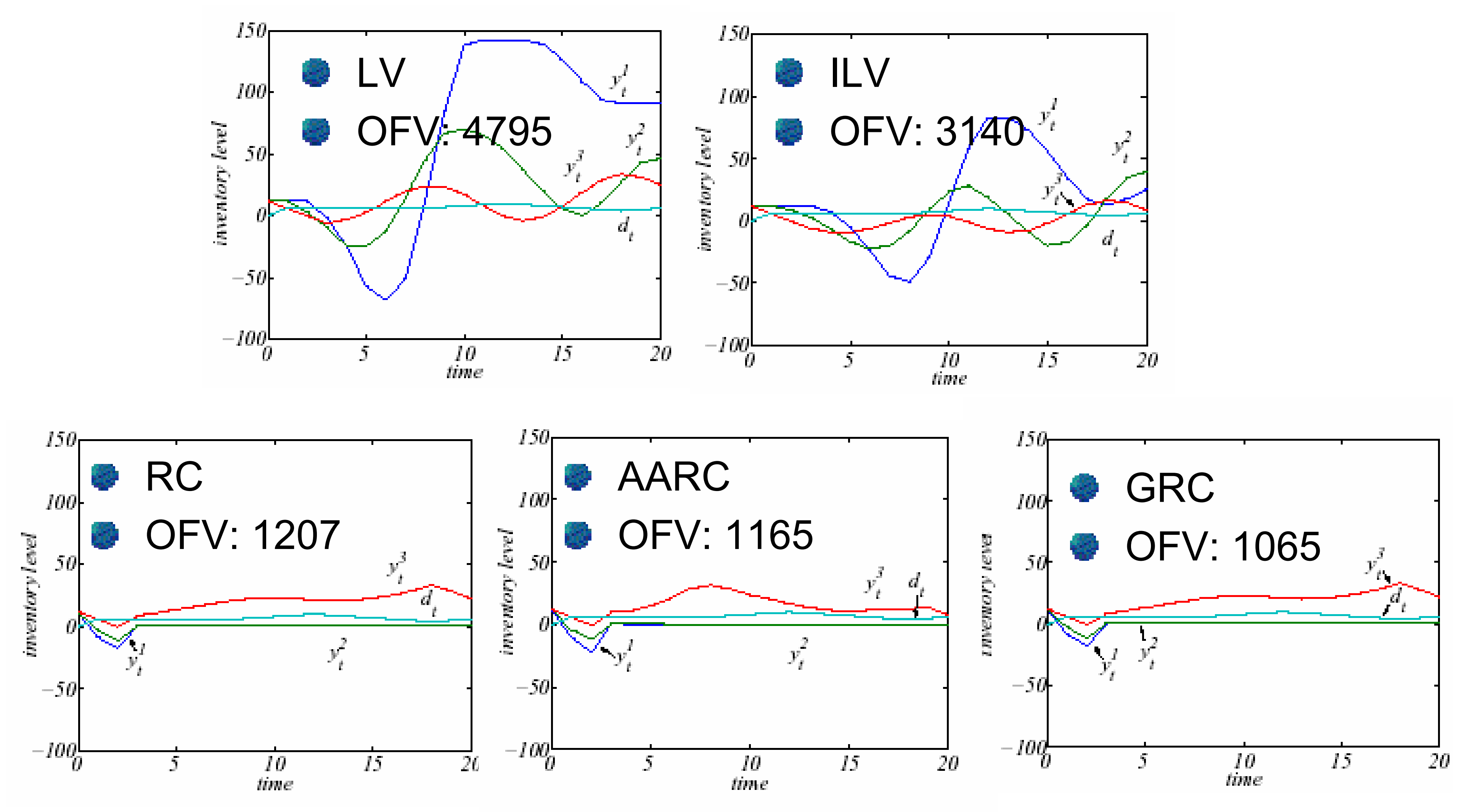
$$v_t^j = \begin{cases} z_0^m - \sum_{t=1}^{t-T^M(m)} d_z & \text{if } j = m \\ z_0^j & j < m \end{cases}$$

The affine control law is here

$$x_t^j = \eta_0^{x,t,j} + \sum_{l=1}^{m} \sum_{\tau=1}^{T} \eta_{\tau l}^{x,t,j} \, v_\tau^l$$

where $\eta_{\tau l}^{x,t,j} = 0 \qquad \forall\, \tau \geq t$ (non anticipativity)

Inventory Behavior – “amplification of oscillation”



Chance Constraints

$$p(w) \equiv \text{Prob}\left\{ w_0 + \sum_{\ell=1}^{d} z_\ell w_\ell \geq 0 \right\} \geq 1 - \epsilon \qquad \text{(C)}$$

- In general, (C) can be difficult to process:
  - The feasible set X of (C) can be nonconvex, which makes it problematic to optimize under the constraint.
  - Even when convex, X can be "computationally intractable":

    Let $z \sim$ Uniform$([0.1]^d)$. In this case, X is convex (Lagoa et al., 2005); however, *unless $P = NP$, there is no algorithm capable to compute $p(w)$ within accuracy δ in time polynomial in the size of the (rational) data w and in $\ln(1/\delta)$* (L. Khachiyan, 1989).

- When (C) is difficult to process "as it is", one can look for a *safe tractable approximation of (C) — a computationally tractable convex set U_ϵ such that* $U_\epsilon \subset X \equiv \{w : p(w) \geq \epsilon\}$.

Robust Optimization to the Rescue of Chance Constraints . . .

Probabilistic Guarantees via RO

$$f_0(x) + \sum_{l=1}^{d} z_l f_l(x) \leq 0\,. \qquad (1)$$

Assumption

$z_1, z_2, \ldots, z_d$ independent rv's

$z_l \sim \mathbf{P}_l \in \mathcal{P}_l$ (compact all prob. dist. in $\mathcal{P}_l$ has common support $= [-1, 1]$).

Definition A vector x satisfying, for a given $0 < z < 1$:

$$\Pr\{f_0(x) + \Sigma z_l f_l(x) \leq 0\} \geq 1 - \epsilon \quad \text{(chance constraint)} \qquad (2)$$

provides a *safe approximation* of (1).

Challenge Find uncertainty set for z, U_ϵ s.t. the Robust Counterpart of (1):

$$f_0(x) + \Sigma z_l f_l(x) \leq 0,\ \forall\, z \in U_\epsilon \qquad (3)$$

is a safe approximation of (1), i.e., every x satisfying (3) satisfies the CC (2).

Theorem

$$
\begin{aligned}
& U_\epsilon = B \cap (M + E_\epsilon) \\
& B = \{u \in I\!R^d \mid \|u\|_\infty \leq 1\} \\
\text{where} \quad & M = \{u | \mu_l^- \leq u_l \leq \mu_l^+,\ l = 1, \ldots, d\} \qquad (4) \\
& E = \{u | \Sigma u_l^2 / \sigma_l^2 \leq 2 \log(1/\epsilon)\}
\end{aligned}
$$

μ_l^-, μ_l^+ and σ_l are such that

$$
A_l(y) \leq \max(\mu_l^- y, \mu_l^+ y) + \frac{\sigma_l^2}{2} y_l^2, \quad \forall\, l = 1, \ldots, d
$$

where

$$
A_l(y) = \max_{P_l \in \mathcal{P}_l} \log \left(\int \exp(ys) dP_l(s) \right) .
$$

Ambiguity set $\mathcal{P}$

Ambiguity set $\mathcal{P}$ should be such that it is possible to obtain good, computationally tractable upper bounds on

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$$

Most frequently, $\mathcal{P}$ consists of $\mathbb{P}$ with known:

- mean
- (co)variance matrix
- possibly, higher order moment information

Ambiguity set $\mathcal{P}$

Ambiguity set $\mathcal{P}$ should be such that it is possible to obtain good, computationally tractable upper bounds on

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(\mathbf{x}, \mathbf{z})$$

Most frequently, $\mathcal{P}$ consists of $\mathbb{P}$ with known:

- mean
- (co)variance matrix
- possibly, higher order moment information

Major works: Scarf (1958), Dupačová (1977), Birge and Wets (1987), Birge and Dulá (1991), Gallego (1992), Gallego, Ryan & Simchi-Levi (2001), Delage and Ye (2010), Wiesemann et al. (2014) and many others...

Forgotten result of Ben-Tal and Hochman (1972)

An exact upper bound when the dispersion measure is the mean absolute deviation (MAD).

Theorem

Assume that a one-dimensional random variable z has support included in $[a, b]$ and its mean and mean absolute deviation are μ and d:

$$\mathcal{P} = \{\mathbb{P} : \quad supp(z) \subseteq [a, b], \quad \mathbb{E}_{\mathbb{P}} z = \mu, \quad \mathbb{E}_{\mathbb{P}} |z - \mu| = d\} .$$

Then, for any convex function $g : \mathbb{R} \to \mathbb{R}$ it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} g(z) = p_1 g(a) + p_2 g(\mu) + p_3 g(b),$$

where $p_1 = \frac{d}{2(\mu - a)}, \quad p_3 = \frac{d}{2(b - \mu)}, \quad p_2 = 1 - p_1 - p_3.$

Generalization to multiple dimensions

The result of Ben-Tal and Hochman (1972) generalizes to multidimensional $\mathbf{z}$ with independent components.

$$\mathcal{P} = \{\mathbb{P} : \text{supp}(z_i) \subseteq [a_i, b_i], \quad \mathbb{E}_{\mathbb{P}} z_i = \mu_i, \quad \mathbb{E}_{\mathbb{P}} |z_i - \mu_i| = d_i, \quad z_i \perp z_j\} \, .$$

Independence implies that the worst-case distribution is a product of the per-component worst-case distributions.

For each convex $g(\cdot)$ it holds that

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} g(\mathbf{z}) = \sum_{\alpha \in \{1,2,3\}^{n_{\mathbf{z}}}} \left(\prod_{i=1}^{n_{\mathbf{z}}} p^i_{\alpha_i} \right) g(\tau^1_{\alpha_1}, \ldots, \tau^{n_{\mathbf{z}}}_{\alpha_{n_{\mathbf{z}}}})$$

where $p^i_{\alpha_i}$ and $\tau^i_{\alpha_i}$ depend only on a_i, b_i, μ_i, and d_i (not on $g(\cdot)$).

Safe approximations

As such, (WC-CC) is intractable and we need a *safe approximation* - a computationally tractable set $\mathcal{S}$ of deterministic constraints such that

$$\mathbf{x} \text{ feasible for } \mathcal{S} \Rightarrow \mathbf{x} \text{ feasible for (WC-CC)}$$

How to construct safe approximations?
The crucial step is a construction of an upper bound on the moment generating function (MGF) of $\mathbf{z}$ (Ben-Tal et al. (2009)):

$$\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T\mathbf{z}).$$

.

Recall: For each convex $g(\cdot)$ it holds that

$$\sup_{\mathbb{P}\in\mathcal{P}} E_{\mathbb{P}} g(z) = \sum_{\alpha\in\{1.2.3\}^n} \left(\prod_{i=1}^{n} p^i_{\alpha_i}\right) g\left(\tau^1_{\alpha^1}, \ldots, \tau^{n_z}_{\alpha_{n_z}}\right)$$

where $p^i_{\alpha_i}$ and $\tau^i_{\alpha_i}$ depend only on a_i, b_i, μ_i, and d_i (not on $g(\cdot)$).

This formula has 3^n terms!

However:

$$\sup_{\mathbb{P}\in P} \log\left(E_{\mathbb{P}} \exp(w^T z)\right) = \sup_{\mathbb{P}\in\mathcal{P}} \log\left(E_{\mathbb{P}}\left(e^{w_1 z_1+\cdots+w_n z_n}\right)\right)$$

$$= \sup_{\mathbb{P}\in\mathcal{P}} \log\left(E_{\mathbb{P}} \prod_{i=1}^{n} e^{w_i z_i}\right) = \text{ due to } z_i\text{'s being independent}$$

$$= \sup_{\mathbb{P}\in\mathcal{P}} \log\left(\prod_{i=1}^{n} E\, e^{w_i z_i}\right) = \sup_{\mathbb{P}\in\mathcal{P}} \sum_{i=1}^{n} \left(\log E\, e^{w_i z_i}\right).$$

So here we need to apply the (B-H) upper (lower) bound separately to each on the n one-variable convex functions $E\, e^{w_i z_i}$!

MGF with our distributional assumptions

We know exactly the worst-case value of the MGF (not just an upper bound):

$$
\begin{aligned}
\sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(\mathbf{w}^T\mathbf{z}) &= \prod_{i=1}^{n_z} \sup_{\mathbb{P}\in\mathcal{P}} \mathbb{E}_{\mathbb{P}} \exp(w_i z_i) \\
&= \prod_{i=1}^{n_z} \left(\frac{d}{2}\exp(-w_i) + 1 - d + \frac{d}{2}\exp(w_i) \right) \\
&= \prod_{i=1}^{n_z} (d\cosh(w_i) + 1 - d)
\end{aligned}
$$

Using this fact, we are able to construct three safe approximations of increasing tightness and increasing complexity.

An example of a safe approximation

Theorem

Let

$$[\mathbf{a}(\mathbf{z}); b(\mathbf{z})] = [\mathbf{a}^0; b^0] + \sum_{i=1}^{n_{\mathbf{z}}} z_i [\mathbf{a}^i; b^i].$$

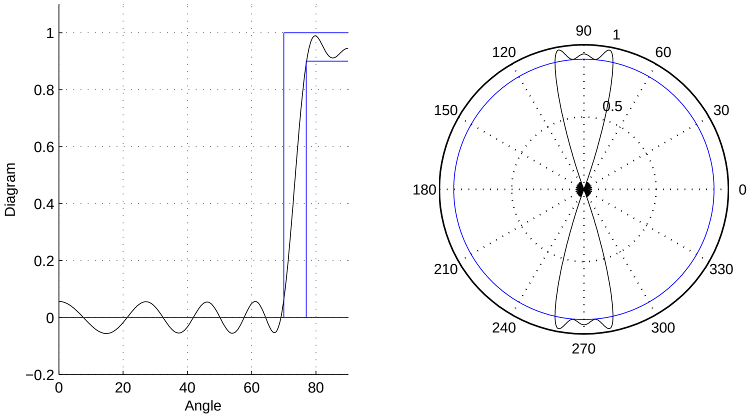
If there exists $\alpha > 0$ *such that* $(\mathbf{x}, \alpha)$ *satisfies the constraint*

$$(\mathbf{a}^0)^T \mathbf{x} - b_0 + \alpha \log \left(\sum_{i=1}^{n_{\mathbf{z}}} \left(d_i \cosh \left(\frac{(\mathbf{a}^i)^T \mathbf{x} - b^i}{\alpha} \right) + 1 - d_i \right) \right) + \alpha \log(1/\epsilon) \leq 0,$$

then $\mathbf{x}$ *satisfies the (WC-CC):* $\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{P} \left(\mathbf{a}^T(\mathbf{z}) \mathbf{x} > b(\mathbf{z}) \right) \leq \epsilon$*.*

The approximating constraint is convex in $(\mathbf{x}, \alpha)$!

Desired diagram graphically



Optimization problem to be solved

$$
\begin{aligned}
\min \quad & \tau \\
\text{s.t.} \quad & -\tau \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq \tau, \quad 0 \leq \phi \leq 70^\circ \\
& -1 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 70^\circ \leq \phi \leq 77^\circ \\
& 0.9 \leq \sum_{i=1}^{n} x_i D_i(\phi) \leq 1, \quad 77^\circ \leq \phi \leq 90^\circ
\end{aligned}
$$

Typically, decisions x_i suffer from implementation error z_i:

$$x_i \mapsto \widetilde{x}_i = (1 + z_i)x_i$$

We want each constraint to hold with probability at least $1 - \epsilon$!

Implementation error

Typically, decisions x_i suffer from implementation error z_i:

$$x_i \mapsto \widetilde{x}_i = (1 + \rho z_i) x_i$$

We want each constraint to hold with probability at least $1 - \epsilon$ for all $\mathbb{P} \in \mathcal{P}$, for example:

$$\mathbb{P}\left(\sum_{i=1}^{n} x_i (1 + \rho z_i) D_i(\phi) \leq 1 \right) \geq 1 - \epsilon, \quad 77^\circ < \phi \leq 90^\circ, \quad \forall \mathbb{P} \in \mathcal{P}$$

Two solutions:

- nominal: no implementation error
- robust: $\rho = 0.001$ and $\epsilon = 0.001$.

Ben-Tal and Hochman (1972)

Assume $f : \mathbb{R} \mapsto \mathbb{R}$ is convex and z follows distribution $\mathbb{P}$ belonging to ambiguity set $\mathcal{P}$ such that

$$\mathcal{P} = \left\{ \mathbb{P} : \text{supp}(z) = [a^-, a^+], \quad \mathbb{E}_{\mathbb{P}} z = \mu, \quad \mathbb{E}_{\mathbb{P}} |z - \mu| = d \right\}$$

Then it holds that:

$$\sup_{\mathbb{P} \in \mathcal{P}} \mathbb{E}_{\mathbb{P}} f(z_i) = p_1 f(a^-) + p_2 f(\mu) + p_3 f(a^+)$$

where $p_1 = \frac{d}{2(a^+ - \mu)}, \quad p_3 = \frac{d}{2(\mu - a^-)}, \quad p_2 = 1 - p_1 - p_3.$

Using this to the MGF problem with our assumptions on $z_1, \ldots, z_n$ we have

$$\sup_{\mathbb{P} \in \mathcal{P}_{(\mu,d)}} \mathbb{E}_{\mathbb{P}} \exp\left(\mathbf{w}^T \mathbf{z}\right) = \prod_{i=1}^{n} \left(d_i \cosh(w_i) + 1 - d_i\right)$$

How does it apply to antenna implementation error

We need $\mu_i^-, \mu_i^+, \sigma_i$ such that:

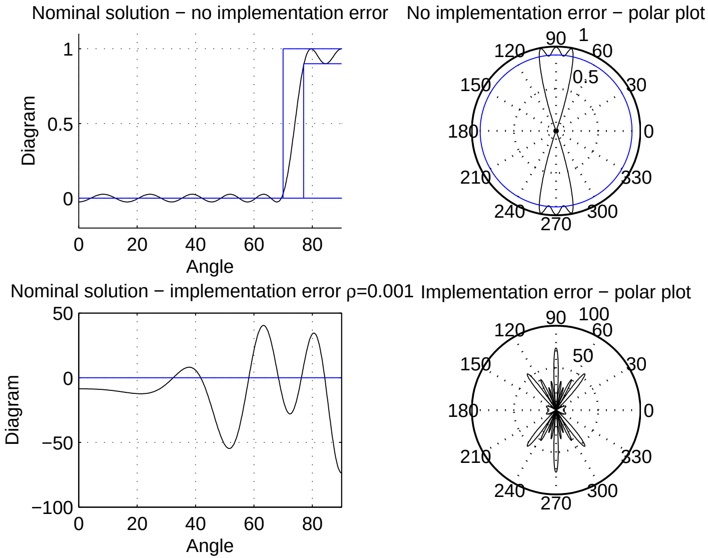
$$d_i \cosh(t) + 1 - d_i \leq \exp\left(\max\{\mu_i^+ t, \mu_i^- t\} + \frac{1}{2}\sigma_i^2 t\right)$$

for all $t \in \mathbb{R}$.
We easily find the right values $\mu_i^- = \mu_i^+ = 0$ and

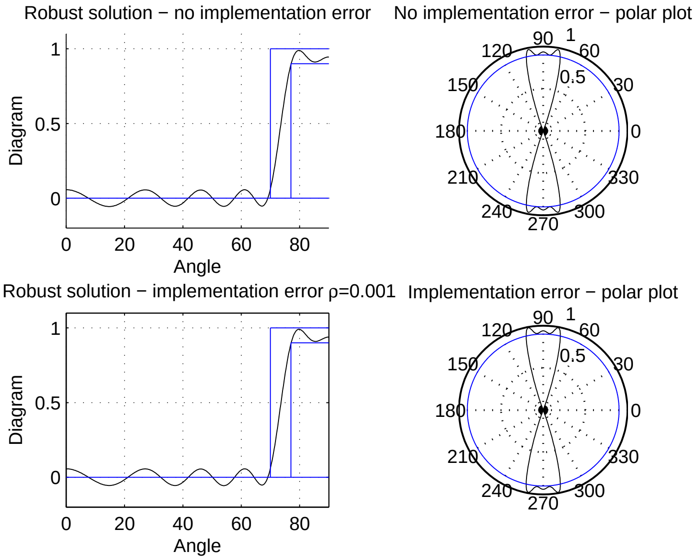
$$\sigma_i = \sup_{t \in \mathbb{R}} \sqrt{\frac{2\log\left(d_i \cosh(t) + 1 - d_i\right)}{t^2}}$$

to satisfy this requirement.

Nominal solution - dream and reality

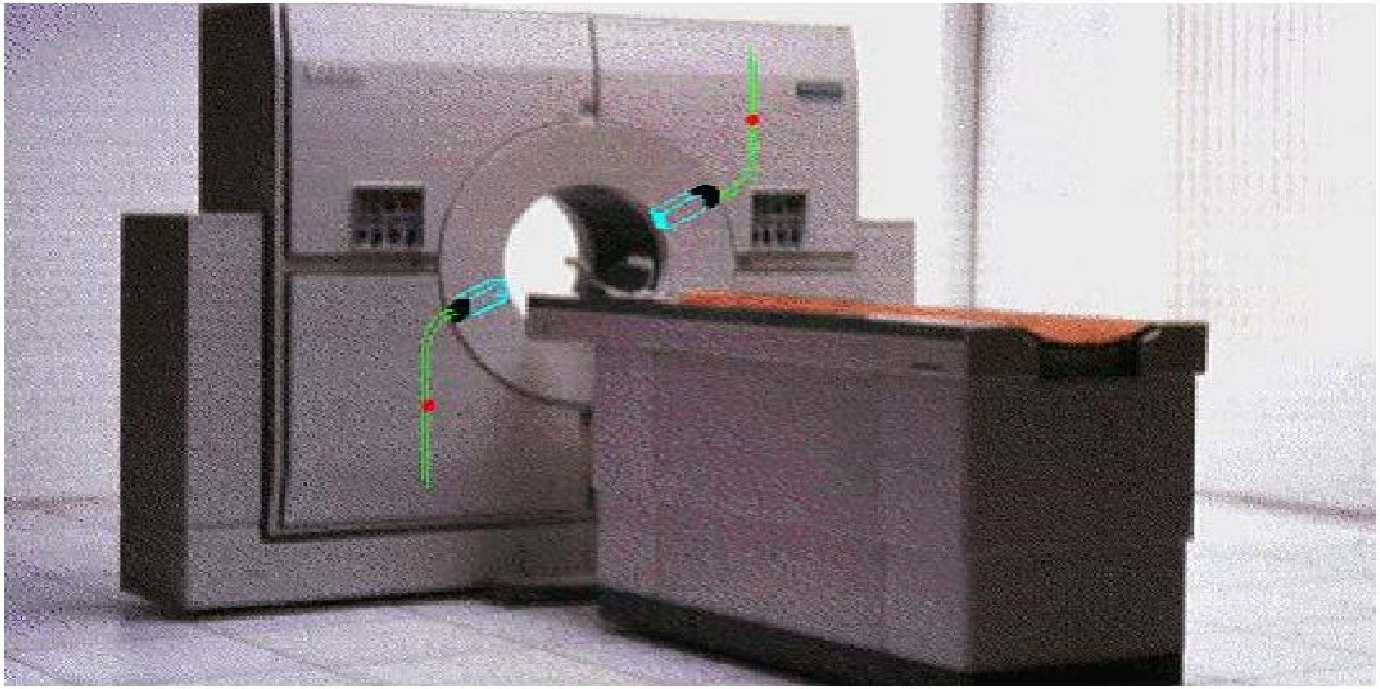


Robust solution - dream and reality



Convex Optimization in the Service of Medicine

Application Example: 3D Imaging in Positron Emission Tomography



♣ PET is a powerful non-invasive medical diagnostic imaging technique for measuring the metabolic activity of cells in human body.
PET imaging is unique in that it shows the chemical functioning of organs and tissues, not just anatomic structures.

♣ An idealized mathematical model of the PET imaging problem is to recover a 2D (or 3D) density from its Radon transform — collection of integrals of the density along all lines in $\mathbf{R}^2$ (or $\mathbf{R}^3$).

♠ In reality, the Radon transform data registered by PET scanner are incomplete, noisy and discretized, which badly affects the quality of the Inverse Radon Transform imaging.

♣ Applying the Maximum Likelihood method, one ends up with the following convex optimization problem:

$$\min_{\lambda}\left\{f(\lambda) = -\sum_{i=1}^{m} y_i \ln\left(\sum_{j=1}^{m} p_{ij}\lambda_j\right) : \lambda \geq 0, \sum_{i=1}^{n} \lambda_i \leq 1\right\} \qquad \textbf{(PET)}$$

- $\lambda \in \mathbf{R}^n$: discretized tracer's density (design vector)
- $y_i \geq 0$ – # of LORs registered by i-th pair of detectors (data)
- $p_{ij} \geq 0$ – probability for LOR originating from j-th grid point to be registered by i-th pair of detectors (data)

♣ PET Imaging problems are extremely large-scale: in 3D,

- the design dimension n varies from 500,000 to 3,000,000
- the number m of log-terms in the objective varies from 3,000,000 to 25,000,000

Reconstruction Algorithms

♠ **When solving typical** *nonlinear* **convex problems, the "price of accuracy digit" for all known polynomial time algorithms is as large as** $O(n^3)$. **With** $n \sim 10^5$, **this price is by six (!) orders of magnitude larger than the performance of modern computers** (~ 1 **Gfl/sec).**

⇒ **With known polynomial time methods, one cannot solve in a realistic time nonlinear convex problems with tens/hundreds of thousands design variables: just the very first iteration will last forever...**

Example: 3D Positron Emission Tomography Imaging by the "best fitting" IP method:

image resolution	n	CPU time per iteration (performance 1 Gfl/sec)
$64 \times 64 \times 64$	**262,144**	**2,5 hours**
$128 \times 128 \times 128$	**2,097,152**	$>$ **13 days**

MIRROR DESCENT METHOD
(Black-Box setting)

Problem

$$f_* = \min_{x \in X} f(x)$$

X convex compact set

f convex Lipschitz continuous on X:

$$|f(x) - f(y)| \leq L\|x - y\| \qquad \forall\, x, y \in X$$

f is given by a *first-order oracle* — a routine which, given $x \in X$, returns the value $f(x)$ and a subgradient $f'(x)$.

Iteration t given x_t, $f(x_t)$, $f'(x_t)$

$$x_{t+1} = \arg\min_{y \in X} \left\{ \ell_{x_t}(y) + \frac{1}{\gamma_t} \omega_{x_t}(y) \right\}$$

$$\begin{array}{rcll} \ell_{x_t}(y) & = & f(x_t) + (y - x_t)^T f'(x_t) & \text{linearization of } f \\ \omega_{x_t}(y) & = & \text{"distance"}(y, x_t) & \text{localizer} \\ 1/\gamma_t & = & \text{penalty parameter} & \end{array}$$

Classical ***Gradient Projection Method***:

$$\omega_x(y) = \|x - y\|_2^2$$

- Best method for $X = \{x \mid \|x\|_2 \leq 1\}$.

For X being a *simplex* :

$$X = \{x \mid \Sigma x_i = 1, \quad x \geq 0\}$$

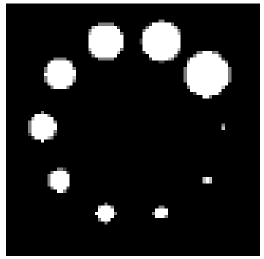
(x_i are "probabilities") a classical distance function in statistical information theory, etc. is the *relative entropy*

$$\omega_x(y) = \Sigma y_i \log(y_i/x_i)\,.$$

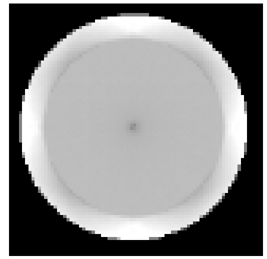
With this choice, we get the MD algorithms for

$$\min_{x \in X} f(x) \qquad X = \text{simplex.}$$

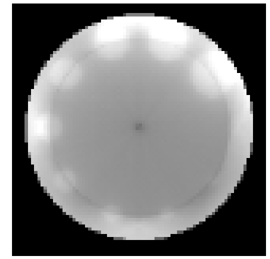
♣ **Experiment 1: noiseless measurements (brighter image correspond to higher tracer's density):**



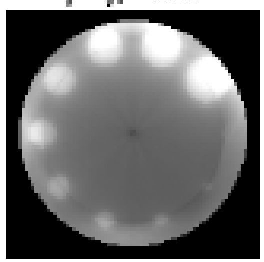
True image: 10 "hot spots"
$f = f_* = 2.817$



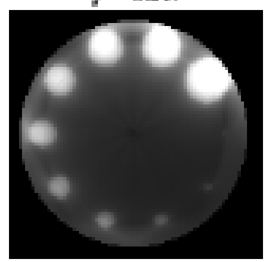
$x^1 = n^{-1}(1, \ldots, 1)^T$
$f = 3.247$



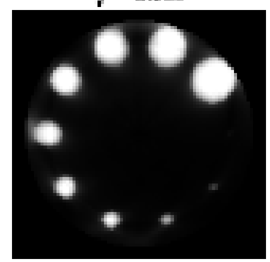
x^2 – some traces of 8 spots
$f = 3.185$



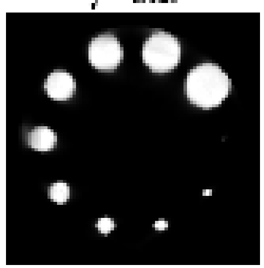
x^3 – traces of 8 spots
$f = 3.126$



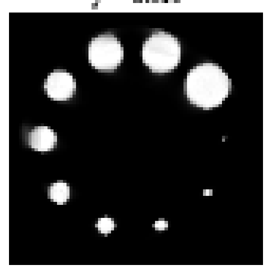
x^4 – some traces of 9-th spot
$f = 3.016$



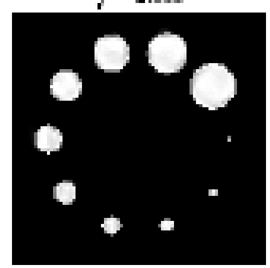
x^8 – 10-th spot still missing...
$f = 2.869$



x^{24} – trace of 10-th spot
$f = 2.828$

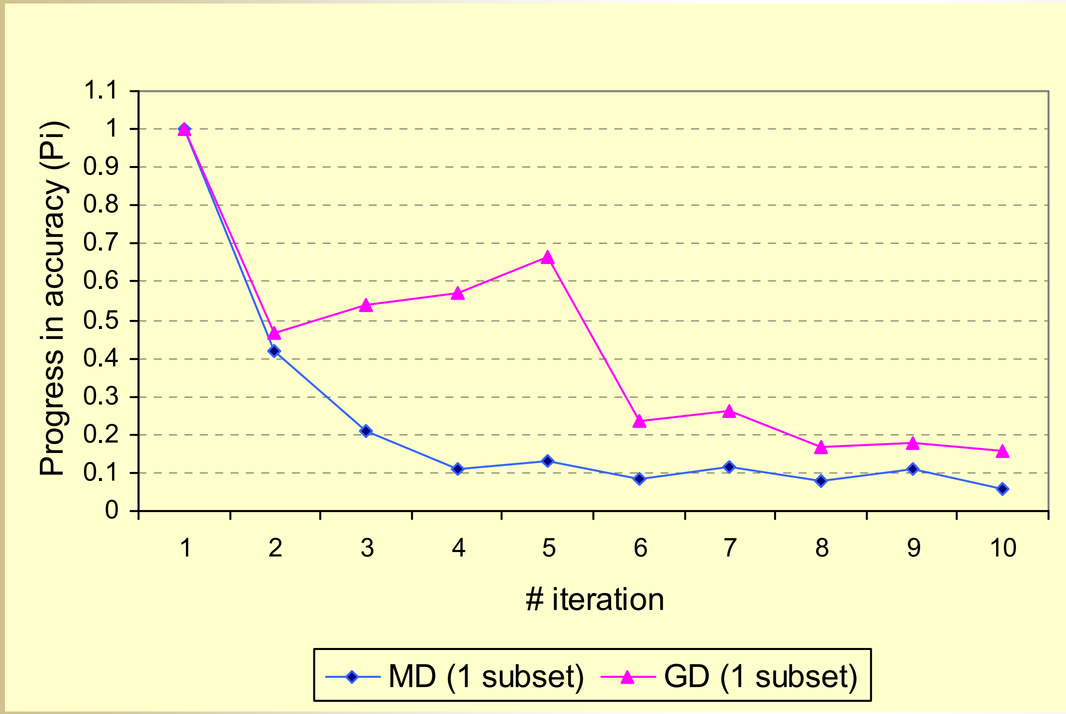


x^{27} – all 10 spots in place
$f = 2.823$

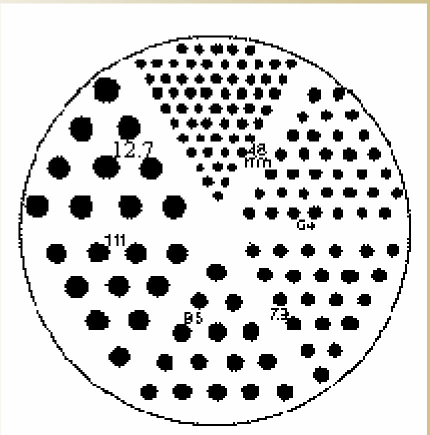


x^{31} – that is it...
$f = 2.818$

Jaszczak Phantom (*n*=515,871)

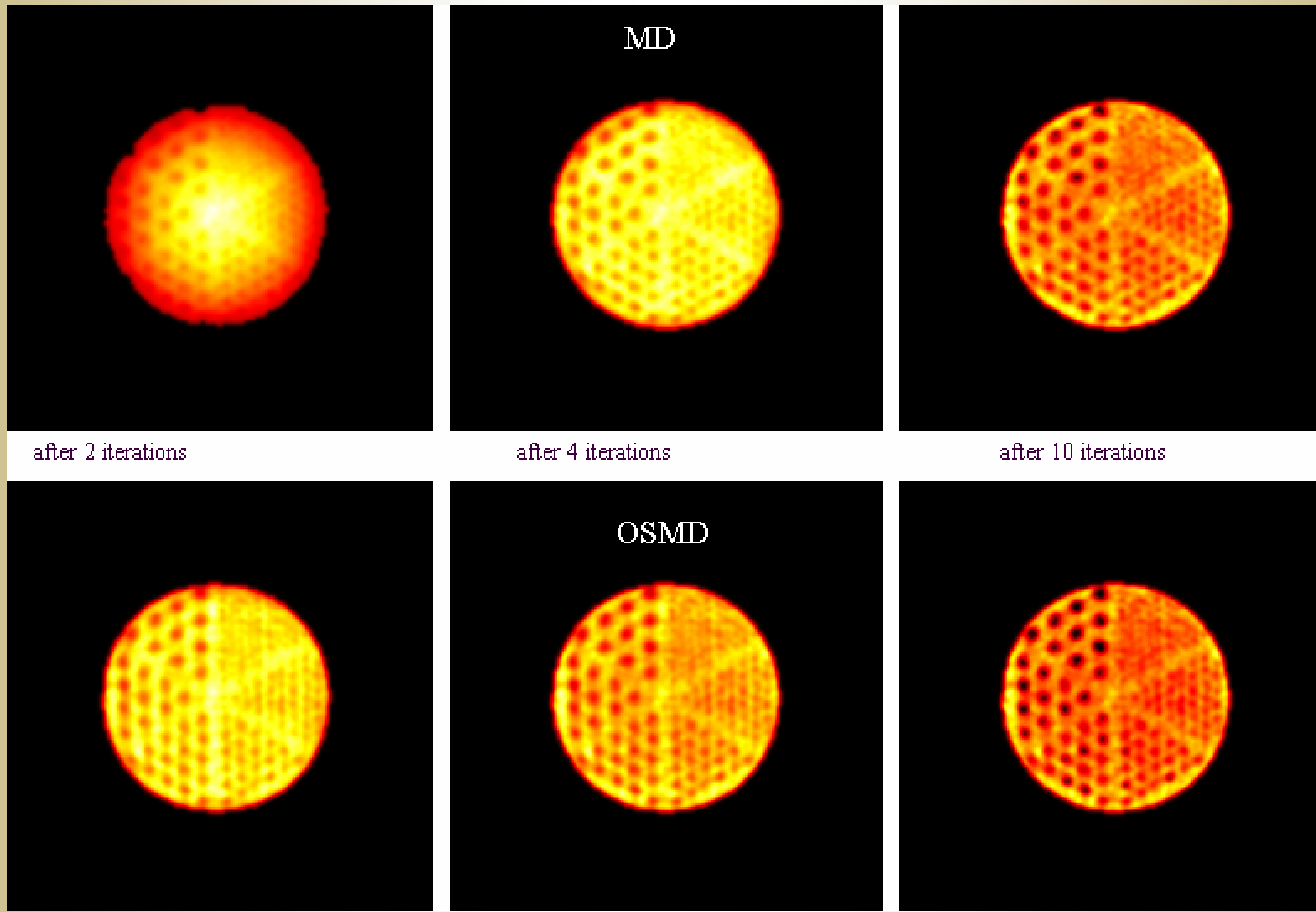


$$P_i = \frac{f_i - f_*}{f_1 - f_*}$$

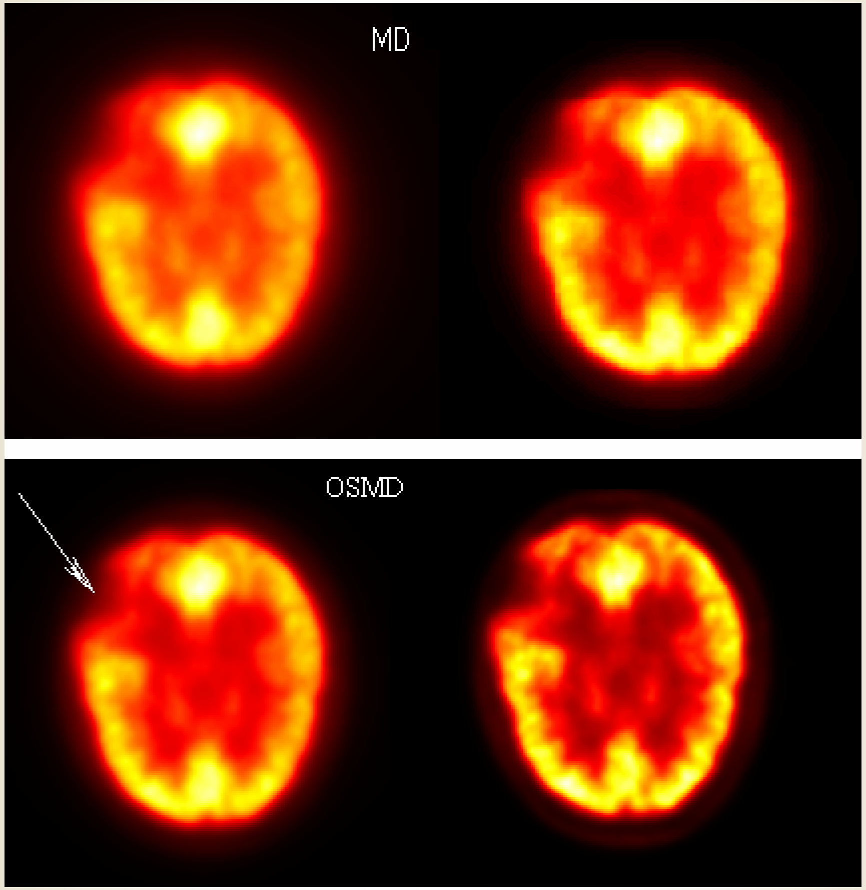


$$f_1 = -5.022\text{e}7;\ \ f_{\text{bst}} = -6.030\text{e}7;\ \ f_* \geq -6.094\text{e}7$$

Jaszczak Phantom (reconstruction by MD *and* OSMD*)*



Brain study -clinical (reconstruction by MD *and* OSMD*)*
GE Advance Tomograph, n=2,763,635, # bins=25,000,000



SVM (Support Vector Machine)

PRIMAL

$$\min_{w,b,\epsilon} \frac{1}{2} \|w\|^2 + c\,\Sigma\,\xi_i$$

$$y_i(w^T x_i + b) \geq 1 - \xi \qquad \forall\, i$$

DUAL

$$\max_{\alpha} \left(\Sigma\alpha_i - \frac{1}{2}\alpha^T Y K Y \alpha \right)$$

$$\alpha \in S_n = \{\alpha \mid 0 \leq \alpha_i \leq c, \quad \Sigma\alpha_i y_i = 0\}$$

$$Y = \text{diag}(y_i \ldots y_n) \quad K_{ij} = K(x_i, x_j)\ [e.g.\ x_i x_j]$$

Uncertainty in kernel matrix

$$K_{ij} = \overline{K}_{ij} + Z_{ij}$$

Chance constraint approach:

$$\text{Dual} \;\Leftrightarrow\; \max_{\alpha,t} \Sigma\alpha_i - \frac{1}{2}\,t$$

$$\alpha \in S_n$$

$$\textbf{Prob}\left(\alpha^T Y \left(\overline{K} + Z\right) Y\alpha \leq t\right) \geq \mathbf{1 - \epsilon}$$

Th. 1 $\quad Z_{ij} \sim N(0, \sigma_{ij})$

CC: $\quad \text{Prob}\left(\alpha^T Y \left(\overline{K} + Z\right) Y \alpha \le t\right) \ge 1 - \epsilon$

holds if

$$\alpha^T Y \overline{K} Y \alpha - \phi^{-1}(\epsilon) \left[\sum_{i,j} \sigma_{ij} \alpha_i^2 \alpha_j^2\right]^{1/2} \le -t$$

Th. 2 $\quad Z_{ij}$ independent r.v.

$$E\, Z_{ij} = 0, \quad \text{supp} Z_{ij} = [a_{ij}, b_{ij}]$$

CC holds if

$$\alpha^T Y \overline{K} Y \alpha + \rho(\epsilon) \left[\sum_{i,j} \beta_{ij} \alpha_i^2 \alpha_j^2\right]^{1/2} \le t$$

$$\rho(\epsilon) = \sqrt{2 \log(1/\epsilon)}$$

$0 \le \beta_{ij}$ simple functions of a_{ij}, b_{ij}.

Denote $\quad Q = Y\, \overline{K}\, Y$

CC dual SVM problem is then approximated by

$$\boxed{\max_{\alpha \in S_n} \left\{ \Sigma \alpha_i - \alpha^T Q \alpha - \rho \left[\sum_{i,j} \beta_{ij} \alpha_i^2 \alpha_j^2 \right]^{1/2} \right\}} \quad (1)$$

$$\Leftrightarrow \quad \min_{\alpha, v} \left\{ \alpha^T Q \alpha + \rho \left[\Sigma \beta_{ij} v_i v_j \right]^{1/2} - \Sigma\, \alpha_i \right\}$$

$$\alpha \in S_n, \quad \boxed{v_i = \alpha_i^2}$$

We can use $\quad v_i \geq \alpha_i^2 \quad$ (recall $\beta_{ij} \geq 0$)

$$\min_{\alpha \in S_n} \left[\alpha^T Q \alpha - \Sigma \alpha_i + \underbrace{\min_{v} \left\{ \rho (v^T B v)^{1/2} \mid v_i \geq \alpha_i^2 \right\}}_{v^*} \right]$$

Result:

$$v^* = \max_{\mu \geq 0} \left\{ \Sigma \mu_i \alpha_i^2 \mid \mu^T B^{-1} \mu \geq \rho \right\}$$

So, finally (1) becomes

$$\boxed{\min_{\alpha \in S_n} \max_{\mu \geq 0} \left\{ \Sigma \mu_i \alpha_i^2 + \alpha^T Q \alpha - \Sigma \alpha_i \mid \mu^T B^{-1} \mu \leq \rho \right\}} \quad (2)$$

convex concave saddle function

Saddle function problem

(P)

$$\min_{\alpha\in\mathcal{A}}\max_{\mu\in\mathcal{M}}\left\{K(\alpha,\mu)=\Sigma\mu_i\alpha_i^2+\alpha^TQ\alpha-\Sigma\alpha_i\right\}=\text{SadVal}$$

$$\mathcal{A}=\{\alpha\mid 0\le\alpha_i\le c,\quad \Sigma\alpha_i y_i=0\}$$

$$\mathcal{M}=\left\{\mu\ge 0\mid \mu^TB^{-1}\mu\le\rho\right\}$$

PROPERTIES

$K(\alpha,\mu)$ is *strongly convex* in α, *linear* in μ

constraint sets are convex, compact

∇K is Lipschitz continuous on $\mathcal{A}\times\mathcal{M}$

Define $\overline{\phi}(\alpha)=\max_{\mu\in\mathcal{M}}K(\alpha,\mu)$ $\underline{\phi}(\mu)=\min_{\alpha\in\mathcal{A}}K(\alpha,\mu)$

Error at a feasible point (α,μ):

$$\epsilon_{Sad}=\overline{\phi}(\alpha)-\underline{\phi}(\mu)\,.\quad(\alpha,\mu\text{ optimal if }\epsilon_{sad}=0)$$

Algorithm [Juditski, Nemirovski, 2010].

Solves SadVal problems under the above properties with ϵ_{sad} at iteration t being of order $O(1/t^2)$.

For our specific problem (P), the main computational effort is to solve at iteration t the following generic problems:

$$(i)\ \min_{\mu \geq 0} \left\{ \frac{1}{2}(\mu - \mu_t)^T B^{-1}(\mu - \mu_t) - p^T x \mid \mu^T B^{-1} \mu \leq \rho \right\}$$

$$(ii)\ \min_{\alpha} \left\{ \frac{1}{2}(\alpha - \alpha_t)^T (\alpha - \alpha_t) - q_t^T \alpha \mid 0 \leq \alpha_i \leq c,\ \Sigma \alpha_i y_i = 1 \right\}$$

Solution of *(i)* is:

$$\mu_{t+1} = \rho \frac{B b_t}{(b_t^T B b_t)^{1/2}}, \quad (b_t = p + B^{-1} \mu_t)$$

Solution of *(ii)* is:

$$\forall i : (\alpha_{t+1})_i = \begin{cases} 0 & \text{if} \quad (\alpha_t)_i + (q_t)_i + \lambda^* y_i \leq 0 \\ c & \text{if} \quad (\alpha_t)_i + (q_t)_i + \lambda^* y_i \geq c \\ (\alpha_t)_i + (q_t)_i + \lambda y_i & \text{otherwise} \end{cases}$$

where λ^* is root of

$$\Sigma \alpha_{t+1}(\lambda) y_i = 0$$