ORDER-VALUE OPTIMIZATION AND NEW APPLICATIONS

José Mario Martínez
www.ime.unicamp.br/∼martinez

UNICAMP, Brazil

August 2, 2011

Collaborators

Roberto Andreani (Applied Math - UNICAMP)
Leandro Martínez (Chemistry - UNICAMP)
Flávio Yano (Itaú Bank)
Mário Salvatierra (Fed. Univ. Amazonas)
Giovane César (Applied Math - UNICAMP)
Roberto Marcondes (Computer Science - USP)
Paulo J. Silva (Computer Science - USP)
Cibele Dunder (Itaú Bank)
Luís Felipe Bueno (Applied Math - UNICAMP)
Lucas Garcia Pedroso (Applied Math - UNICAMP)
Maria Aparecida Diniz (Applied Math - UNICAMP)
Ernesto Birgin (Computer Science - USP)

Outline

- Introduce Order-Value Optimization problems
- Review of Algorithms and Convergence results
- Applications

Order-Value Optimization (OVO) Problems

$$\text{Let } f_i : \Omega \subset \mathbb{R}^n \to \mathbb{R}, \quad i = 1, \cdots, m$$
$$J \subset \{1, \cdots, m\}$$

For all $x \in \Omega$, we define $i_1(x), i_2(x), \ldots, i_m(x)$ by

$$f_{i_1(x)}(x) \leq f_{i_2(x)}(x) \leq \cdots \leq f_{i_m(x)}(x)$$

The OVO problem is:

$$\underset{x \in \Omega}{\text{minimize}} \quad \sum_{j \in J} f_{i_j(x)}(x)$$

Examples

$$J = \{m\} \longrightarrow \min_{x\in\Omega} \max\{f_1(x), \cdots, f_m(x)\}$$

$$J = \{1\} \longrightarrow \min_{x\in\Omega} \min\{f_1(x), \cdots, f_m(x)\}$$

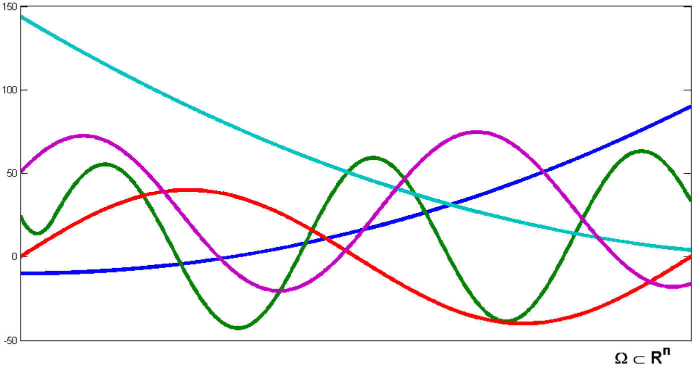
$$J = \{p\} \longrightarrow \min_{x\in\Omega} f_{i_p(x)}(x) \quad \text{(VaR-Like)}$$

$$J = \{1, \cdots, p\} \longrightarrow \min_{x\in\Omega} \sum_{j=1}^{p} f_{i_j(x)}(x) \quad \text{(LOVO)}$$

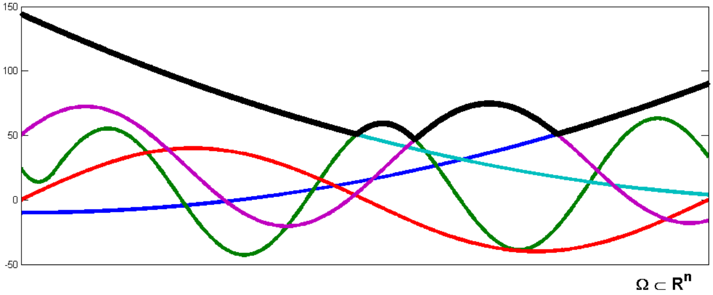
$$J = \{p+1, \cdots, m\} \longrightarrow \min_{x\in\Omega} \sum_{j=p+1}^{m} f_{i_j(x)}(x) \quad \text{(CVaR-Like)}$$

$$J = \{q+1, \cdots, p\} \longrightarrow \min_{x\in\Omega} \sum_{j=q+1}^{p} f_{i_j(x)}(x)$$

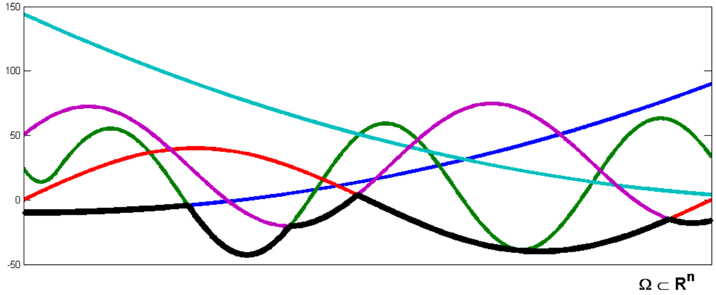
Non-smoothness and Many local minimizers



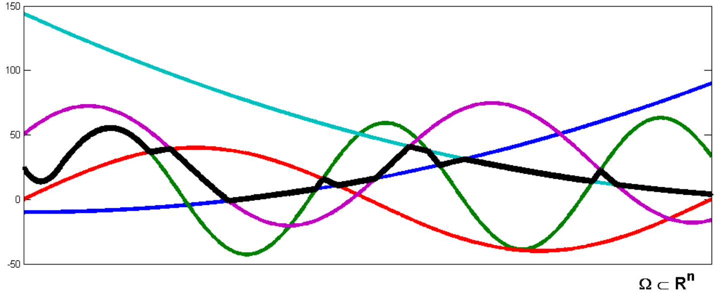
$J = \{m\}$ (Minimax)



$J = \{1\}$ (Minimin)



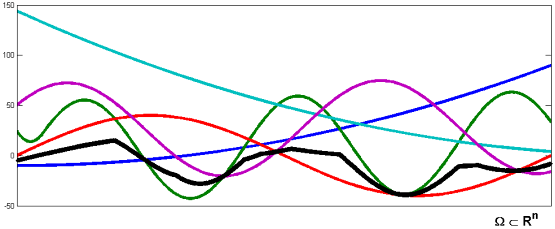
$J = \{p\}$, (VaR-like)



In this example, $p = 3$

$J = \{1, \cdots, p\}$, (LOVO)

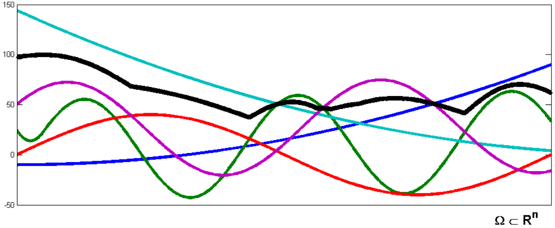
$$\min_{x \in \Omega} \frac{1}{p} \sum_{j=1}^{p} f_{i_j(x)}(x)$$



In this example, $p = 2$

$J = \{p+1, \cdots, m\}$, (CVaR-like)

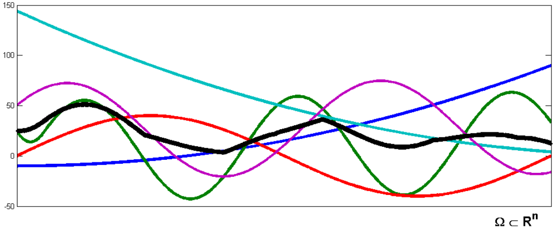
$$\min_{x\in\Omega} \frac{1}{m-p} \sum_{j=p+1}^{m} f_{i_j(x)}(x)$$



In this example, $p = 3$

$J = \{q+1, \cdots, p\}$

$$\min_{x \in \Omega} \frac{1}{p-q} \sum_{j=q+1}^{p} f_{i_j(x)}(x)$$



In this example, $q = 1, p = 4$

Reformulation of CVaR-like

Fact

$$b_1 \le b_2 \le \cdots \le b_m, \qquad p \le m-1$$

$$\Rightarrow b_{p+1} + \cdots + b_m = \underset{\xi \in \mathbb{R}}{\text{Minimum}} \quad (m-p)\xi + \sum_{i=1}^{m} \max\{0, b_i - \xi\}$$

$$\text{Minimizers} = \{\xi \in [b_p, b_{p+1}]\}$$

$$\Downarrow$$

$$\underset{x \in \Omega}{\text{Minimize}} \quad f_{i_{p+1}(x)}(x) + \cdots + f_{i_m(x)}(x)$$

is equivalent to:

$$\underset{x \in \Omega,\ \xi \in \mathbb{R}}{\text{Minimize}} \quad (m-p)\xi + \sum_{i=1}^{m} \max\{0, f_i(x) - \xi\}$$

Reformulation of VaR-like

From the same "fact"

$$\underset{x \in \Omega}{\text{Minimize}} \quad f_{i_p(x)}(x)$$

is equivalent to:

$$\text{Minimize } \xi \ \ (\text{with respect to } x \in \Omega \ \text{ and } \ \xi \in \mathbb{R})$$

subject to

$$\xi \text{ minimizes } (m-p)\xi + \sum_{i=1}^{m} \max\{0, f_i(x) - \xi\} \ \ (\text{ wrt } \xi)$$

Consequence for the Reformulations of CVaR-like and VaR-like

CVaR-like is a Nonlinear Programming problem
Convex if the f_i are convex
Linear-Programming if the f_i are linear
with many inequality constraints

VaR-like is a Bilevel Programming problem
with many Complementarity Constraints
that come from the KKT conditions of the Lower-Level problem

Primal method for minimizing CVaR-like

Consider

$$\underset{x \in \Omega,\ \xi \in \mathbb{R}}{\text{Minimize}} \quad (m-p)\xi + \sum_{i=1}^{m} \max\{0, f_i(x) - \xi\}$$

Use smoothing to deal with of max and ordinary NLP for minimizing on $\Omega \times \mathbb{R}$.

Primal Method for Minimizing VaR-like

Given the current point $x^k \in \Omega$ (convex) take a sufficient descent direction d^k for all j such that

$$f_{i_p(x^k)}(x^k) - \epsilon \leq f_j(x^k) \leq f_{i_p(x^k)}(x^k) + \epsilon$$

Line-search along $d^k \rightarrow x^{k+1} = x^k + \alpha_k d^k$

Global convergence to (ϵ)stationary points

- Local
- Superlinear
- Quadratic
- Convex subproblems (linear or quadratic constraints)

Risk Minimization

- m scenarios
- $f_i(x)$ =predicted loss caused by decision $x \in \Omega$ under scenario i
- $f_{i_p(x)}(x)$ = VaR associated with x
- $\frac{1}{m-p} \sum_{j=p+1}^{m} f_{i_j(x)}(x)$ = CVaR associated with x
- $\underset{x \in \Omega}{\text{minimize}} \quad f_{i_p(x)}(x) \quad \equiv \quad$ minimize VaR
- $\underset{x \in \Omega}{\text{minimize}} \quad \sum_{j=p+1}^{m} f_{i_j(x)}(x) \quad \equiv \quad$ minimize CVaR

Low Order-Value Optimization (LOVO)

Define, as always, $i_1(x), \ldots, i_m(x)$ by:

$$f_{i_1(x)}(x) \leq \cdots \leq f_{i_m(x)}(x); \qquad p \leq m$$

Then, the LOVO problem is:

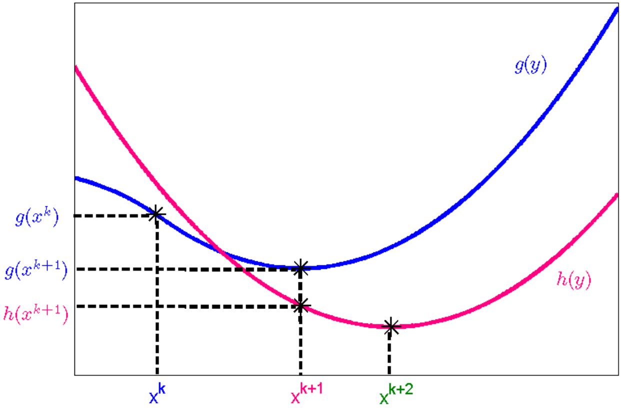
$$\underset{x \in \Omega}{\text{minimize}} \quad \sum_{j=1}^{p} f_{i_j(x)}(x)$$

Fact

$$\sum_{j=1}^{p} f_{i_j(x)}(y) \leq \sum_{j=1}^{p} f_{i_j(x)}(x) \Rightarrow \sum_{j=1}^{p} f_{i_j(y)}(y) \leq \sum_{j=1}^{p} f_{i_j(x)}(x)$$

$$\Downarrow$$

In order to decrease the LOVO function we may "fix" $(i_1(x), \cdots, i_p(x))$ and "minimize" $\sum_{j=1}^{p} f_{i_j(x)}(y)$ with respect to y.



$$g(y) = \sum_{j=1}^{p} f_{i_j(x^k)}(y) \qquad h(y) = \sum_{j=1}^{p} f_{i_j(x^{k+1})}(y)$$

Methods for Unconstrained LOVO problems

<u>Line-Search</u>

At iteration k

Find a sufficient descent direction for $\sum_{j=1}^{p} f_{i_j(x^k)}(x)$

take

$$\sum_{j=1}^{p} f_{i_j(x^k+\alpha_k d)}(x^k+\alpha_k d) \underset{\text{suf}}{<} \sum_{j=1}^{p} f_{i_j(x^k)}(x^k)$$

Global Convergence to points x^* such that

$$\nabla \sum_{j=1}^{p} f_{i_j(x^*)}(x^*) = 0$$

Trust-Region methods for LOVO

Typical iteration

Given x^k, the trust region defined by Δ and a quadratic approximation of $\sum_{j=1}^{p} f_{i_j(x^k)}(x)$:

Minimize the quadratic approximation on the trust region Δ

- If the reduction of $\sum_{j=1}^{p} f_{i_j(x^k)}(x)$ is sufficiently large with respect to the reduction of the quadratic approximation (Ared ≥ 0.1 Pred) accept the solution of the trust region subproblem as x^{k+1}.
- Otherwise, reduce Δ.

Convergence of trust-region methods for LOVO

At every limit point x^*,

$$\nabla \sum_{j=1}^{p} f_{i_j(x^*)}(x^*) = 0.$$

Using the true Hessian to define the quadratic approximation:

$$\nabla^2 \sum_{j=1}^{p} f_{i_j(x^*)}(x^*) \geq 0$$

Local convergence: quadratic

Fitting with LOVO

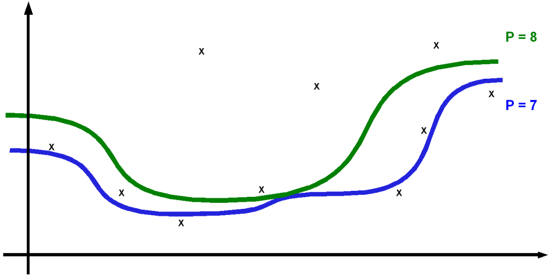
Observations

t	y
t_1	y_1
$\vdots$	$\vdots$
t_m	y_m

Model

$$y_j \approx M(x, t_j)$$

$$f_j(x) = [y_j - M(x, t_j)]^2$$



Constrained LOVO problems

$$\text{Minimize} \sum_{j=1}^{p} f_{i_j(x)}(x)$$

subject to

$$h(x) = 0, g(x) \leq 0.$$

Augmented Lagrangian (PHR-Like) (Code Algencan in www.ime.usp.br/∼egbirgin/tango)

$$\text{Minimize approx} \sum_{j=1}^{p} f_{i_j(x)}(x) + \frac{\rho}{2}\left[\left\| h(x) + \frac{\lambda}{\rho} \right\|^2 + \left\| \left(g(x) + \frac{\mu}{\rho} \right)_+ \right\|^2\right]$$

Update $\lambda, \mu \geq 0, \rho$. $(a_+ = \max\{0, a\})$

Convergence of Algencan-LOVO

Global minimization of subproblems
$\Downarrow$
Global Minimization

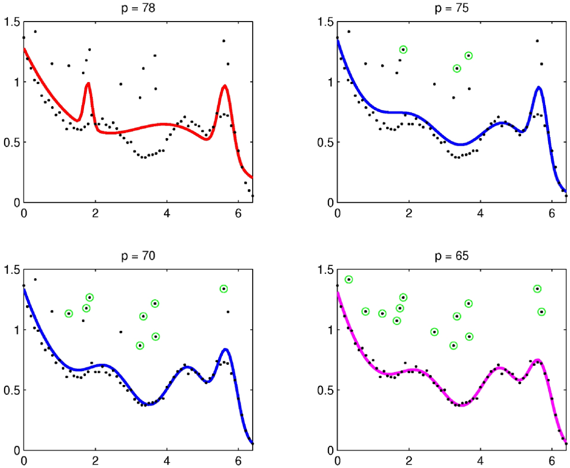
Limit points are either feasible or stationary points of Infeasibility

Feasible limit points that satisfy the CPLD constraint qualification are "KKT"

Boundedness of penalty parameter

Model fitting with Algencan-LOVO

Find the parameters of a Boundary Value Problem fitting a set of data that contains outliers



Fitting Nash-Equilibrium Models

Nash-Equilibrium Model:

Given the parameters $x \in \Omega$, the players $1, 2, \cdots, m$ take , simultaneously, decisions $y_1, \cdots, y_m$.
Player j takes his/her decision minimizing $f_j(x, y_1, \cdots, y_{j-1}, z, y_{j+1}, \cdots, y_m)$ with respect to z.

Fitting Nash-Equilibrium Models

Nash-Equilibrium Model:

Given the parameters $x \in \Omega$, the players $1, 2, \cdots, m$ take , simultaneously, decisions $y_1, \cdots, y_m$.
Player j takes his/her decision minimizing $f_j(x, y_1, \cdots, y_{j-1}, z, y_{j+1}, \cdots, y_m)$ with respect to z.

Inverse Nash-Equilibrium:

$\bar{y}_1, \cdots, \bar{y}_m$ are known
Discover the parameters x.

Fitting Nash-Equilibrium Models

Nash-Equilibrium Model:

Given the parameters $x \in \Omega$, the players $1, 2, \cdots, m$ take , simultaneously, decisions $y_1, \cdots, y_m$.
Player j takes his/her decision minimizing $f_j(x, y_1, \cdots, y_{j-1}, z, y_{j+1}, \cdots, y_m)$ with respect to z.

Inverse Nash-Equilibrium:

$\bar{y}_1, \cdots, \bar{y}_m$ are known
Discover the parameters x.

LOVO-Inverse-Nash-Equilibrium:

$\bar{y}_1, \cdots, \bar{y}_m$ are known but only 90% of these observations are reliable.

Fitting Nash-Equilibrium Models

$$\text{Minimize } \sum_{j=1}^{p} (y_{i_j(x,y)} - \bar{y}_{i_j(x,y)})^2$$

subject to

Player j minimizes $f_j(x, y_1, \ldots, y_{j-1}, z, y_{j+1}, \ldots, y_m)$

with respect to z, for all $j = 1, \ldots, m$

Dealing with LOVO constraints

Minimize f(x)

s/t x satisfies at least p constraints of the subset $\begin{cases} g_1(x) \leq 0 \\ \vdots \\ g_m(x) \leq 0 \end{cases}$

We define $L(x, \mu, \rho) = f(x) + \rho \sum_{j=1}^{p} \left(g_{i_j(x)}(x) + \frac{\mu_{i_j}}{\rho} \right)_+^2$

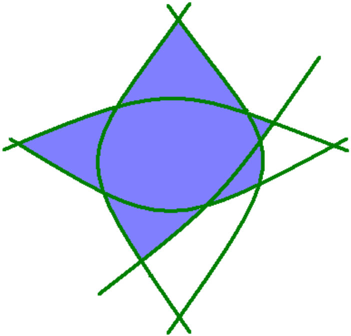
where $\left(g_{i_1(x)}(x) + \frac{\mu_{i_1}}{\rho} \right)_+^2 \leq \cdots \leq \left(g_{i_m(x)}(x) + \frac{\mu_{i_m}}{\rho} \right)_+^2$

minimize $L(x, \mu, \rho)$

update μ, ρ

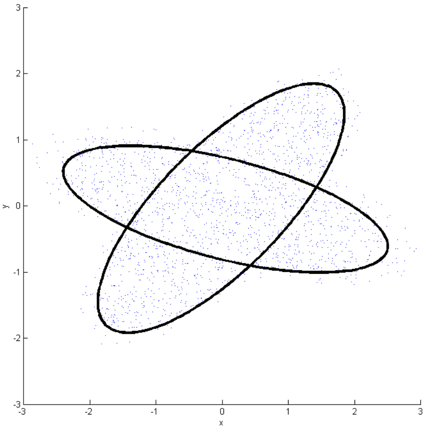
LOVO constraints

Example of Feasible Region for $p = m - 1$



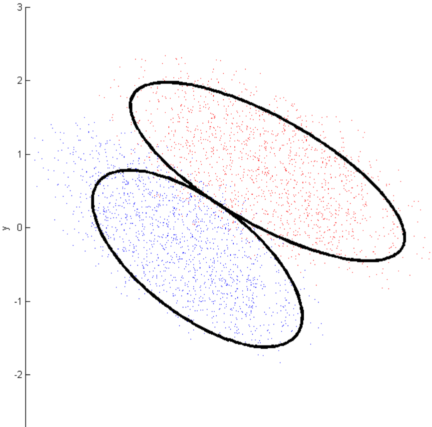
Example of Minimization with LOVO constraints

Find the union of 2 ellipses with smallest area that contain 90% of the points $\{P_1, \cdots, P_m\}$
We have one constraint $P_i \in E_1 \cup E_2$ for each point P_i.



Example of Minimization with LOVO constraints

Given two sets of points A and B, find two ellipses with no intersection area, that contain, respectively, 80% of the points of A and 80% of the points of B.



Topics in computer vision

Curve detection {
- detection of lines
- detection of circles
- detection of ellipses
- others

[Comparisons with HT, Ransac, QMDPE, LKS in Cesar-Andreani-Marcondes-JMM & Silva 2007]

Lines

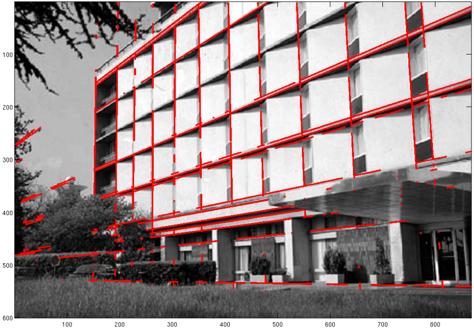
Given $\{(x_1, y_1), \cdots, (x_m, y_m)\}$ find θ and ρ such that

$$\sum_{j=1}^{p} f_{i_j(\theta,\rho)}(\theta, \rho) \text{ is minimal,}$$

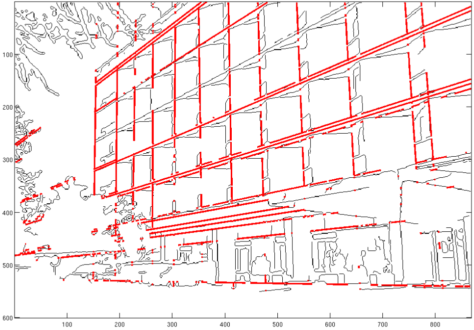
where $f_i(\theta, \rho) = [x_i \cos\theta + y_i \sin\theta - \rho]^2 \; i = 1, \cdots, m,$

and $f_{i_1(\theta,\rho)}(\theta, \rho) \leq \cdots \leq f_{i_m(\theta,\rho)}(\theta, \rho) \quad \forall\theta$ and $\rho > 0$

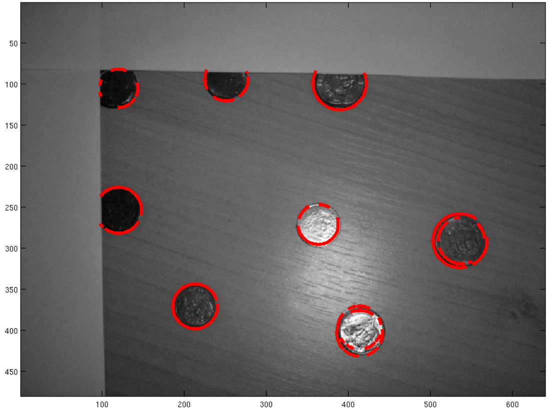
Detection of lines



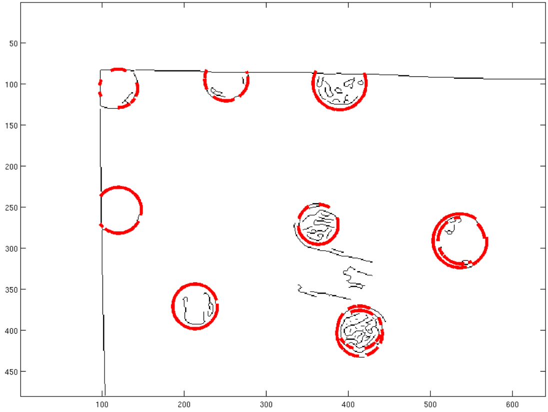
Detection of lines



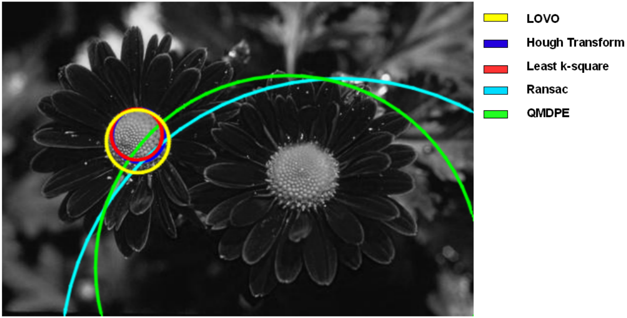
Detection of circles



Detection of circles



Detection of circles



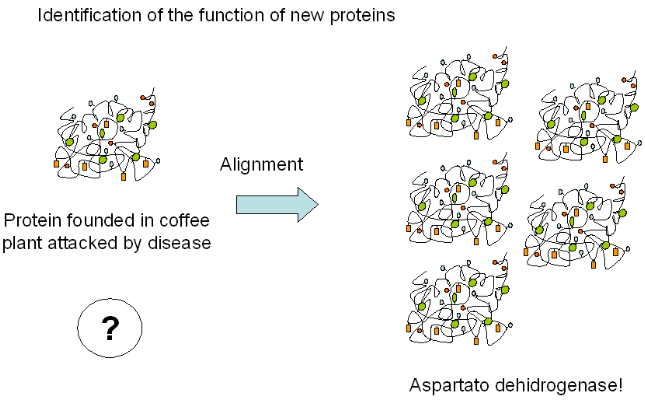
Current work

Tracking Parametric Curves in video sequences

The tracked curve in frame t-1 should be used as initial solution for the LOVO problem to be solved in frame t.

Protein Alignment

Finding common 3D structures of two given proteins.
Evaluating similarity.



Protein Alignment

Data: 3D coordinates of $C - \alpha$ atoms of both proteins.

(Available, for example, in Protein Data Bank PDB)

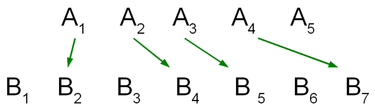
Protein A: $(A_1, \cdots, A_{n_A})$

Protein B: $(B_1, \cdots, B_{n_B})$

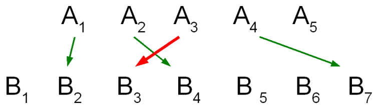
$A_i, B_j \in \mathbb{R}^3$

To each admissible bijection between (a subset of) A and (a subset of) B corresponds a score.

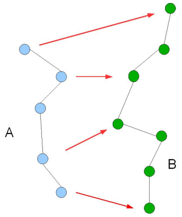
Example admissible bijection Φ



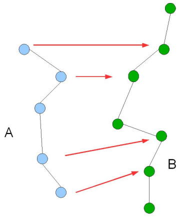
Non-admissible bijection



Given fixed relative positions of A and B, the bijection with best score may be obtained using dynamic programming



Score = 1.31



Score = 13.42

Changing the relative positions, both "best bijection" and score change.

$-f_i(x) =$ score associated with
"bijection i" under the
movement defined by x.

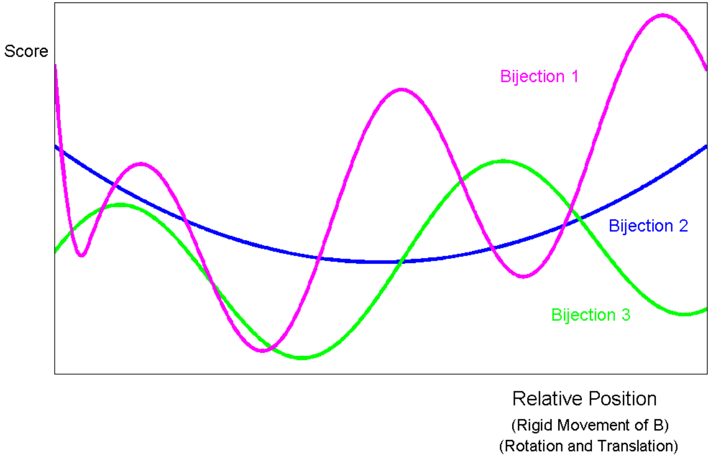
$$\Downarrow$$

$$\text{Minimize } f_{i_1(x)}(x) \text{ (LOVO, } p = 1)$$

Objective of the alignment:

maximize the score, with respect to bijection and relative position.

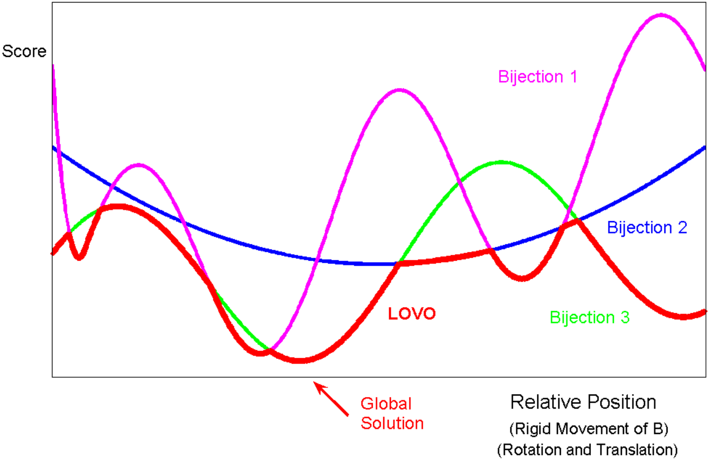
Modelling as LOVO:



Objective of the alignment:

maximize the score, with respect to bijection and relative position.

Modelling as LOVO:



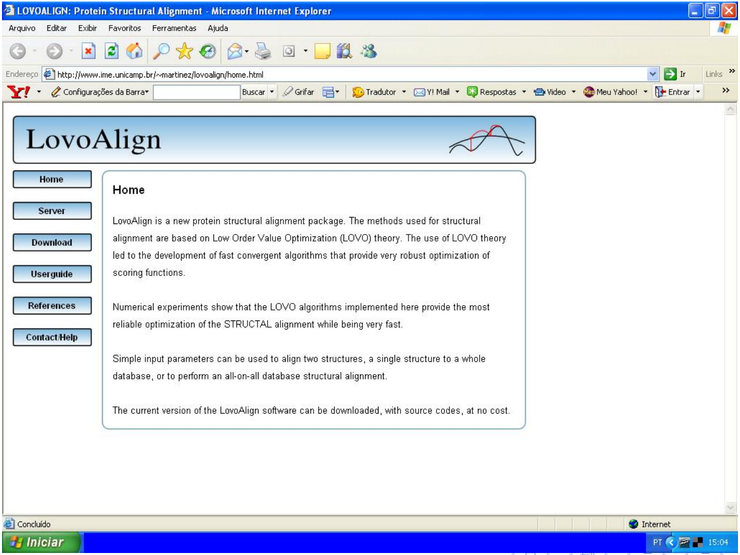
LOVOALIGN PACKAGE

www.ime.unicamp.br/~martinez/lovoalign

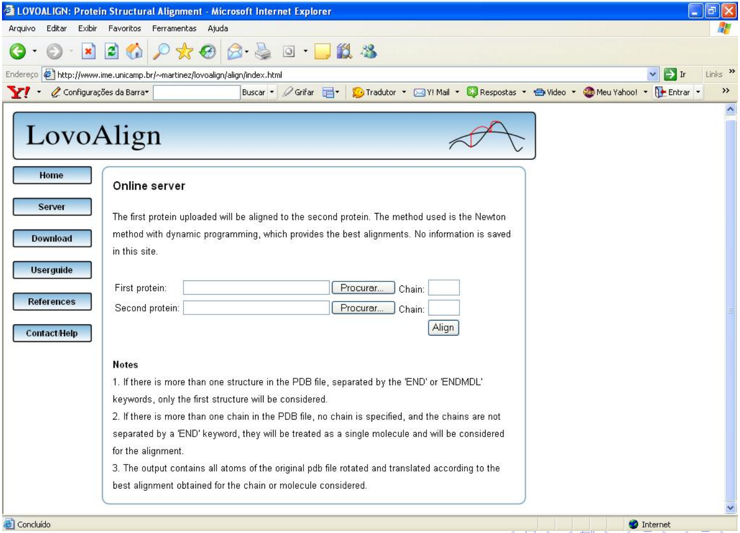
On-line alignment

of proteins

www.ime.unicamp.br/∼martinez/lovoalign



www.ime.unicamp.br/∼martinez/lovoalign



Conclusion

We presented:

- The general form of Order-Value functions and OVO problems
- Particular cases: VaR-like, CVaR-like, LOVO, ...
- Discussion of nonsmoothness and many local minimizers
- Nonlinear Programming Reformulations
- Primal (trust-region and line-search) methods for unconstrained OVO problems with and without smoothing
- Using LOVO for Nonlinear Regression with Outliers
- Constrained LOVO problems
- Convergence of Algorithms for Constrained LOVO
- Application to Nash-Equilibrium fitting
- LOVO constraints
- Computer Vision
- Protein Alignment

Some references

1. R. Andreani, J. M. Martínez, L. Martínez and F. Yano. Continuous Optimization Methods for Structure Alignments. To appear in *Mathematical Programming*.
2. R. Andreani, J. M. Martínez, M. Salvatierra and F. Yano. Quasi-Newton methods for order-value optimization and value-at-risk calculations. *Pacific Journal of Optimization* 2, pp. 11-33 (2006).
3. R. Andreani, C. Dunder and J. M. Martínez. Nonlinear-Programming Reformulation of the Order-Value Optimization problem. *Mathematical Methods of Operations Research* 61, pp. 365-384 (2005).
4. R. Andreani, C. Dunder and J. M. Martínez. Order-Value Optimization: formulation and solution by means of a primal Cauchy method. *Mathematical Methods of Operations Research* 58, pp. 387-399 (2003).