Canonical Duality: Theory, Methods and Applications



SAPIENZA
UNIVERSITÀ DI ROMA

VITTORIO LATORRE

OVERVIEW

- Canonical Duality Theory
- The interior Point Method
- Sensor Network Localization Problem
- Numerical Results

DUALITY

- Duality is a theory that transforms a whole optimization problem into another
- The Dual problem is generally simpler to solve than the original and shows properties of the phenomenon that cannot be understood by just looking at the original formulation
- After the dual problem is solved, it is easy to find the solution to the original problem
- Generally a minimization problem is transformed into a maximization problem

DUALITY

- Duality was first developed in linear optimization
- It is possible to create a new formulation of the original problem that has the following proprieties:
  - Weak duality: the value of the objective function in the primal problem is always bigger than the value of the objective function of the dual problem
  - Strong duality: the value of the objective functions at the optimum are the same
- These conditions can be used to find an optimal solution of certain classes of algorithms
  - Reduce the duality gap to zero
- In problems with a big number of variables and few constrains the dual could be easier to solve

DUALITY

- Optimality conditions also for semidefinite programming
- Optimality conditions for constrained optimization
  - Lagrange Theory and Karush-Kuhn-Tucker conditions
- Useful results for convex programming
- Few useful results in non-convex optimization
  - The presence of several local minima creates a “duality gap” that cannot be filled
- Duality theory proved useful in all of those fields where there is no duality gap
  - Without duality gap it is possible to find global optimality conditions

FENCHEL CONJUGATE

- In unconstrained optimization, given the convex function $f(x)$ we can define

$$f^*(\sigma) = \sup_{x \in R^n} \{x\sigma - f(x)\}$$

- As the Legendre-Fenchel conjugate of $f(x)$
- This function is such that:
  - the conjugate of the conjugate is the original function:

$$f^{**}(\sigma) = f(x)$$

  - $f^*(\sigma) \geq x\sigma - f(x)$
  - There are the following relations between primal and dual variables:

$$x = \nabla f^*(\sigma) \quad \sigma = \nabla f(x)$$

  - If $\bar{x} = \arg\min f(x)$ and $\bar{\sigma} = \arg\min f^*(\sigma)$ we have:

$$f^*(\bar{\sigma}) = \bar{x}\bar{\sigma} - f(\bar{x})$$

CANONICAL DUALITY

- Canonical Duality a generalization to Fenchel duality that can also be applied to non convex functions
- It is composed of
  - A Canonical Dual Transformation
  - A complementary-dual principle
  - A triality theory
- Suppose that we want to solve the following general non-convex problem

$$\min_{x \in R^n} \left\{ P(x) = W(x) + \frac{1}{2} x^T A x - F^T x \right\}$$

CANONICAL DUALITY

- Key Idea:
  - Find a certain geometrical non-linear measure

$$\xi = \Lambda(x) : R^n \to \mathrm{E}_a \subseteq R^m$$

  - Such that

$$W(x) = V(\Lambda(x)) \quad V(\xi) : \mathrm{E}_a \to \subseteq R,$$

  - With $V(\xi)$ convex
- Consequently the primal problem can be rewritten as

$$\min_{x \in R^n} \left\{ P(x) = V(\Lambda(x)) + \frac{1}{2} x^T A x - F^T x \right\}$$

LEGENDRE TRANSFORMATION

- If the function $V(\xi)$ is convex it is possible to apply a Legendre transformation, finding its convex Legendre conjugate $V^*(\sigma)$ *of* $V(\xi)$, which has the following property:

$$V^*(\sigma) = \sup_{\xi} \left\{ \langle \xi, \sigma \rangle - V(\xi) \right\}$$

- Where σ is the dual variable, of the same dimension of ξ defined as:

$$\sigma = \nabla_{\xi} V(\xi) : E \to S_a \subseteq R^m$$

- And the following relations hold for both the primal and dual variables in stationary points:

$$\sigma = \nabla_{\xi} V(\xi) \Leftrightarrow \xi = \nabla_{\sigma} V(\sigma)$$

Total Complementarity Function

- By using the Legendre transformation it is possible to substitute $V(\xi) = \langle \xi, \sigma \rangle - V^*(\sigma)$ in the primal problem and obtain:
$$\Xi(x,\sigma) = \langle \Lambda(x), \sigma \rangle - V^*(\sigma) + \frac{1}{2} x^T A x - F^T x$$
- That is the total complementarity function in the primal and dual variables
- The total complementarity function is strictly related to the Lagrange function
  - If the measure Λ is chosen linear, the total complementarity corresponds to the Lagrange function

CANONICAL DUALITY

- In many real world application the non linear operator $\Lambda(x)$ is quadratic:

$$\Lambda(x) = \left\{ \frac{1}{2} x^T C_k x - x^T b_k \right\}^m : R^n \to E_a \subset R^m$$

- And the Total complementarity function can be written in the following way:

$$\Xi(x, \sigma) = \frac{1}{2} x^T G(\sigma)\, x - F(\sigma)^T x - V^*(\sigma)$$

$$G(\sigma) = A + \sum_{k=1}^{m} C_k \sigma_k, \quad F(\sigma) = c + \sum_{k=1}^{m} \sigma_k b_k$$

- That is quadratic in the primal variable

Dual Problem

- The dual is obtained by exploiting the stationary conditions in the primal variable

$$\nabla_x \Xi(x,\sigma) = 0_n \Rightarrow x = G(\sigma)^{-1} F(\sigma),$$

- By substituting this newfound value in the total complementarity function, we obtain:

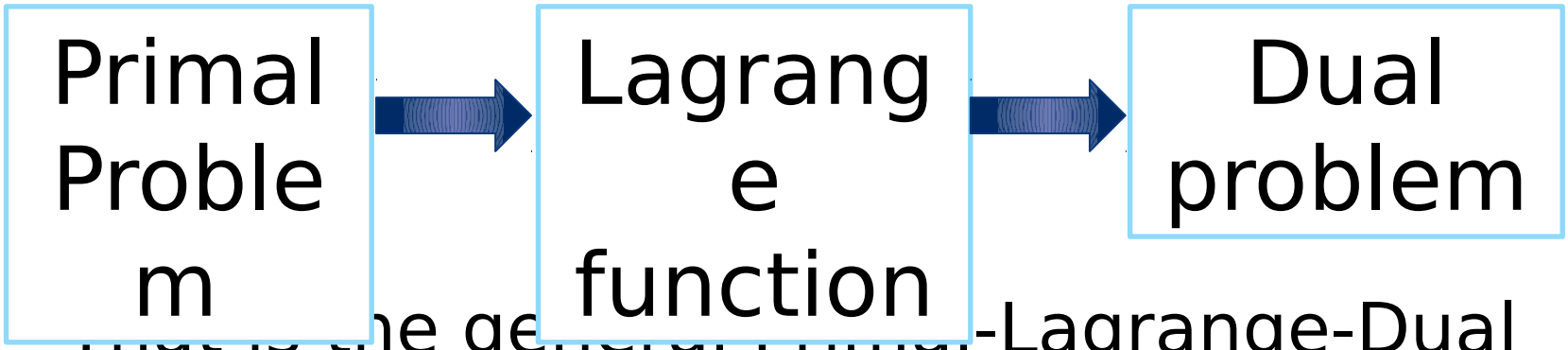
$$P^*(\sigma) = -\frac{1}{2} F(\sigma)^T G(\sigma)^{-1} F(\sigma) - V^*(\sigma).$$

RELATIONS

- We have:

$$P(x) \Rightarrow \Xi(x, \sigma) \Rightarrow P^d(\sigma)$$

- That is similar to the following framework:



- That is the general Primal-Lagrange-Dual framework often encountered in optimization.
- The dual also has some important proprieties that can be used for finding the global minimum of the problem efficiently

Total Complementarity Principle

- Theorem 1: (Total Complementarity Principle) the function $Pd(\sigma)$ is canonically dual to $P(x)$ in the sense that if σ^* is a critical point of $Pd(\sigma)$ then the vector:

$$x^* = G\left(\sigma^*\right)^{-1} F\left(\sigma^*\right),$$

- Is a critical point of $P(x)$ and

$$P\left(x^*\right) = \Xi\left(x^*, \sigma^*\right) = P^d\left(\sigma^*\right).$$

- Conversely if x^* id a solution of $P(x)$ it must be in the previous form for a critical point σ^* of $Pd(\sigma)$

Dual Proprieties

- Theorem 1 clearly states that there is no duality gap between the critical points of the primal problem and the critical points of the dual
  - Effectively creating the dual of a non convex problem without any duality gap
  - Once a dual stationary point has been found, the corresponding solution in the primal can be easily found by solving a linear system
- The second propriety gives the global optimality conditions. To this aim we define two sets:

$$S_a^+ = \{\sigma \in S_a \mid G(\sigma) \succeq 0\},$$

$$S_a^- = \{\sigma \in S_a \mid G(\sigma) \prec 0\}$$

Triality Theory

- Given a critical point (x^*, σ^*) of $\Xi(x, \sigma)$, the following three extremality conditions hold:
  - **Global Optimality**: the critical solution x^* is the unique global minimizer of $P(x)$ if and only if σ^* belongs to S^+_a i.e:

$$\min_{x \in R^n} P(x) = P(x^*) = \Xi(x^*, \sigma^*) = P^d(\sigma^*) = \max_{\sigma \in S^+_a} P^d(\sigma)$$

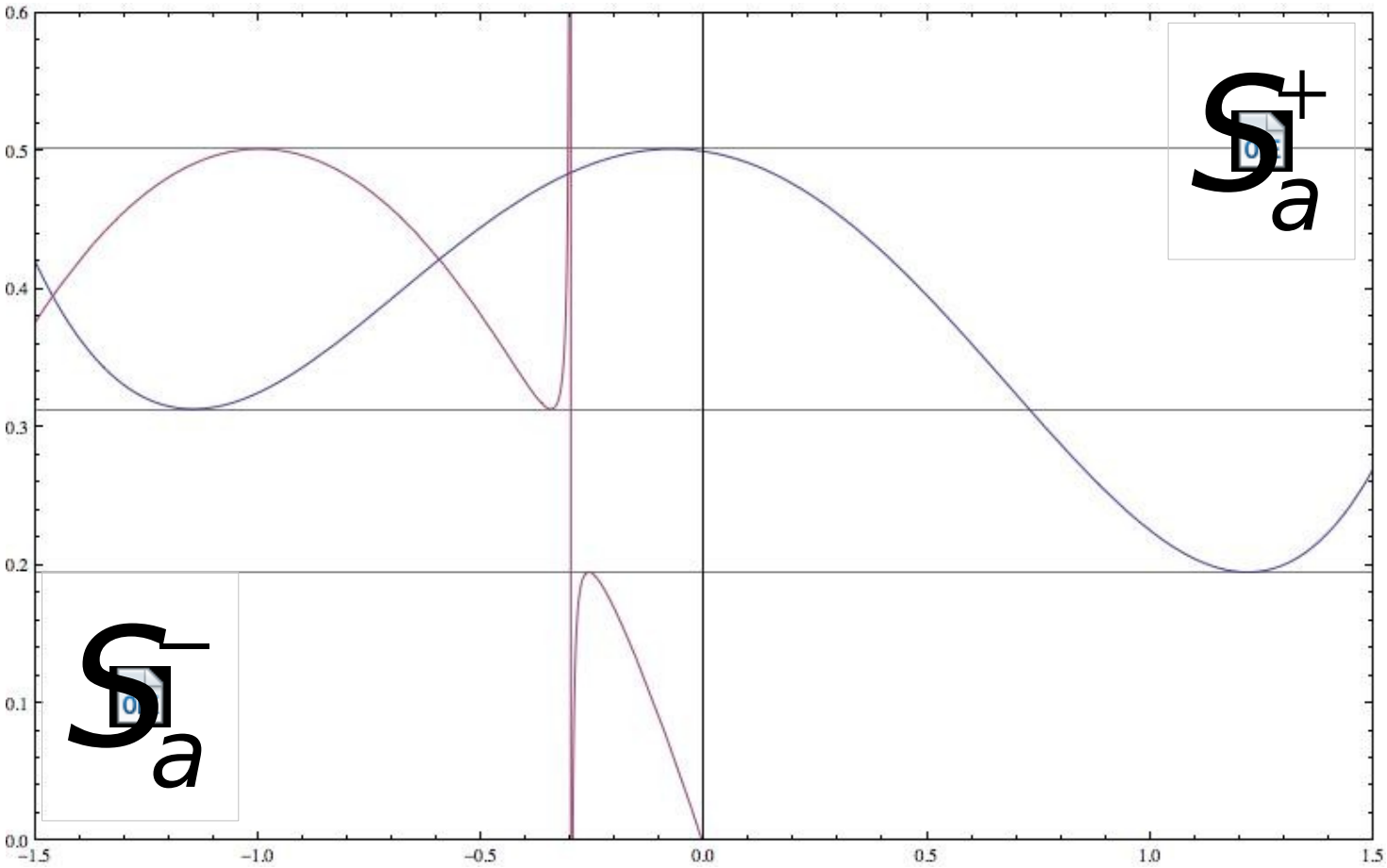
  - **Highest local maximum**: if σ^* belongs to S^+_a then σ^* is a local maximizer of $Pd(\sigma)$ on its neighborhood $S0$ in S^+_a if and only if x^* is a local maximizer of $P(x)$ on its neighborhood X0 in Rn, i.e.

$$\max_{x \in X_0} P(x) = P(x^*) = \Xi(x^*, \sigma^*) = P^d(\sigma^*) = \max_{\sigma \in S^-_a} P^d(\sigma)$$

  - **Highest local minimum**: if σ^* belongs to S^-_a then σ^* is a local minimizer of $Pd(\sigma)$ on its neighborhood $S0$ in S^-_a if and only if x^* is a local minimizer of $P(x)$ on its neighborhood $X0$ in Rn, i.e.:

$$\min_{x \in X_0} P(x) = P(x^*) = \Xi(x^*, \sigma^*) = P^d(\sigma^*) = \min_{\sigma \in S^-_a} P^d(\sigma)$$

Triality Theory



- The red vertical line divides $S+a$ from S-a
- The stationary points have the same value of the objective function

Canonical Duality

- The global optimality conditions give information on where to search for the global optimum of the problem:

$$\max_{\sigma \in R^m} P^d(\sigma) = -\frac{1}{2} F(\sigma)^T G(\sigma)^{-1} F(\sigma) - V^*(\sigma)$$

$$s.t \quad G(\sigma) \succeq 0$$

- Therefore solving this problem will give the corresponding solution to the global minimum in the primal

Canonical Dual Algorithm

- To find the global minimum, we can solve the dual such that $G(\sigma)$ is positive semi-definite.
- This approach has several draws-backs
  - It is necessary to calculate the inverse of the matrix every time the objective function is evaluated
  - Ill-conditioning problems with the inversion
  - Difficulty to recover the primal solution if the algorithm does not converge to a good enough approximation of a stationary point
- Therefore we apply a primal-dual strategy for the solution based on the total complementarity function

Primal-Dual Algorithm

- The problem we want to solve is to find a saddle point of the total complementarity function $\Xi(x,\sigma)$ such that $G(\sigma)$ is positive definite.

$$\min_{x\in R^n}\max_{\sigma\in R^m}\Xi(x,\sigma)=\frac{1}{2}x^T G(\sigma)\,x-F(\sigma)^T x-V^*(\sigma),$$

$$s.t.\quad G(\sigma)\succeq 0$$

- Such problem can be reformulated as the problem of finding the solution of a monotone variational inequality on a convex set

$$H(x,\sigma)=0, G(\sigma)\succeq 0,\quad H(x,\sigma)=\begin{pmatrix}\nabla_x\Xi(x,\sigma)\\ -\nabla_\sigma\Xi(x,\sigma)\end{pmatrix}$$

- To this aim we reformulate these conditions as system of Constrained Equations(CE) and propose an Interior point method specifically designed to solve this system and find a saddle point of $\Xi(x,\sigma)$

Primal-Dual Algorithm

- Therefore we want to solve the following $CE(H,\Omega)$ system:

$$H(x,\sigma) = 0, \sigma \in S_a^+ \equiv \Omega,$$

$$H(x,\sigma) = \begin{pmatrix} \nabla_x \Xi(x,\sigma) \\ -\nabla_\sigma \Xi(x,\sigma) \end{pmatrix}$$

- We use as measure the squared norm of the operator $H(x,\sigma)$ in order to measure the progress of the algorithm
- At every iteration the line-search assures that $G(\sigma)$ is positive semi-definite.

Primal-Dual Algorithm

1. Choose $(x0,\sigma 0),\gamma \in (0,1)$, $\varepsilon>0$, $k=0$
2. If $||H(x,\sigma)||2<\varepsilon$, stop
3. Find a solution dk to the following system of linear equations

$$J H(x^k,\sigma^k)d^k = -H(x^k,\sigma^k).$$

4. Find a step size αk such that

$$(x^k+\alpha^k d_x^k, \sigma^k+\alpha^k d_\sigma^k) \in R^n \times S_a^+$$

$$\left\|H(x^k+\alpha^k d_x^k, \sigma^k+\alpha^k d_\sigma)\right\|^2 \le \left\|H(x^k,\sigma^k)\right\|^2 - \gamma \nabla H\left(x^k,\sigma^k\right)^T d^k$$

5. Set $(xk,\sigma k)=(xk+\alpha k dxk,\sigma k+\alpha k d\sigma k)$, $k=k+1$ and to to step 2

Convergence

- The theoretical convergence of the presented algorithm is still not proved yet.
- There exists a convergence proof based on a new measure of duality that takes inspiration from the Fenchel-Young equation
- But such result must be peer-reviewed before claiming it
- It should be provable that: *if* $\{(xk,\sigma k)\}$ *is the sequence generated by the algorithm, then every limit point of such sequence is a solution of* $CE(H,\Omega)$

Sensor Localization Problem

- A problem that arises in monitoring and controlling applications using wireless sensor networks such as inventory management and gathering environment data.
  - Also related to distance geometry and prediction molecule structures,
  - And graph realization
- It consists in locating N sensors of unknown positions knowing the positions of Na sensors, called anchors, and the sensor-sensor, sensor-anchor distances in a space of dimension dim.

Sensor Localization Problem

- We consider:
  - The dimension of the problem: n=dim*N
  - The variables $x \; in \; RN*dim$ that represent the coordinates of the sensors
  - hij, the distances between the sensors i and j
  - eik, the distances between the sensor i and the anchor k
  - ρ, the radio range. If the distance between a sensor i and a sensor j is bigger than ρ they do not influence each other. Same thing between a sensor and an anchor
- Therefore we introduce the following sets

$$A_h = \{(i,j) : \|x_i - x_j\| \leq \rho, i \neq j\}$$

$$A_e = \{(i,k) : \|x_i - a_k\| \leq \rho\}$$

Sensor Localization Problem

- The Problem can be formulated as a non-linear least squares optimization problem

$$\min_{x \in R^n} P(x) = \frac{1}{2} \sum_{(i,j) \in A_h} \left(\left\| x_i - x_j \right\|^2 - d_{ij}^2 \right)^2 + \frac{1}{2} \sum_{(i,k) \in A_e} \left(\left\| x_i - a_k \right\|^2 - e_{ik}^2 \right)^2$$

  - At the global optimum the value of the solution is zero

- Like showed before, by using the non-linear operators:

$$\xi_{ij}^h = \Lambda_{ij}(x) = \left\| x_i - x_j \right\|^2, \xi_{ik}^e = \Lambda_{ik}(x) = \left\| x_i - a_k \right\|^2,$$

- We can introduce the following quadratic functions:

$$\gamma_h(\xi^h) = \frac{1}{2} \sum_{(i,j) \in A_h} \left(\xi_{ij}^h - d_{ij}^2 \right)^2, \gamma_e(\xi^e) = \frac{1}{2} \sum_{(i,k) \in A_e} \left(\xi_{ik}^e - e_{ik}^2 \right)^2$$

Sensor Localization Problem

- We have the following relations between the primal and the dual variables in the critical points:

$$\varsigma_{ij}^{h} = \xi_{ij}^{h} - h_{ij}^{2}, (i,j) \in A_h,$$

$$\varsigma_{ik}^{e} = \xi_{ik}^{e} - e_{ik}^{2}, (i,k) \in A_k,$$

- And the following Legendre Conjugates:

$$V_h^*(\varsigma^h) = \sum_{(i,j)\in A_h} \frac{1}{2}(\varsigma_{ij}^h)^2 + h_{ij}^2 \varsigma_{ij}^h,$$

$$V_e^*(\varsigma^e) = \frac{1}{2} \sum_{(i,k)\in A_e} \frac{1}{2}(\varsigma_{ik}^e)^2 + e_{ik}^2 \varsigma_{ik}^e,$$

- By setting $\sigma=(\varsigma h,\varsigma e)$ and $V^*(\sigma)=V^*h(\varsigma h)+V^*e(\varsigma e)$ and reordering the terms:

$$\Xi(x,\sigma) = \frac{1}{2} x^T G(\sigma) x - F(\sigma)^T x - V^*(\sigma)$$

$$P^d(\sigma) = -\frac{1}{2} F(\sigma)^T G(\sigma)^{-1} F(\sigma) - V^*(\sigma)$$

Primal-Dual Algorithm

- For sensor network localization it is possible to notice that at the optimum the dual variables must satisfy

$$\zeta_{ij} = \xi_{ij} - h_{ij}^2 = \left\| x_i - x_j \right\|^2 - h_{ij}^2, (i,j) \in A_h,$$

$$\zeta_{ik}^e = \xi_{ik}^e - e_{ik}^2 = \left\| x_i - a_k \right\|^2 - e_{ik}^2, (i,k) \in A_k,$$

- That is at the global optimum for the primal problem the dual variables are all set to zero
- Also by knowing that:

$$R_m^+ \perp S_a^+$$

- It is possible to substitute the semidefinite constrain with simpler non-negativity constraints on the dual variables:

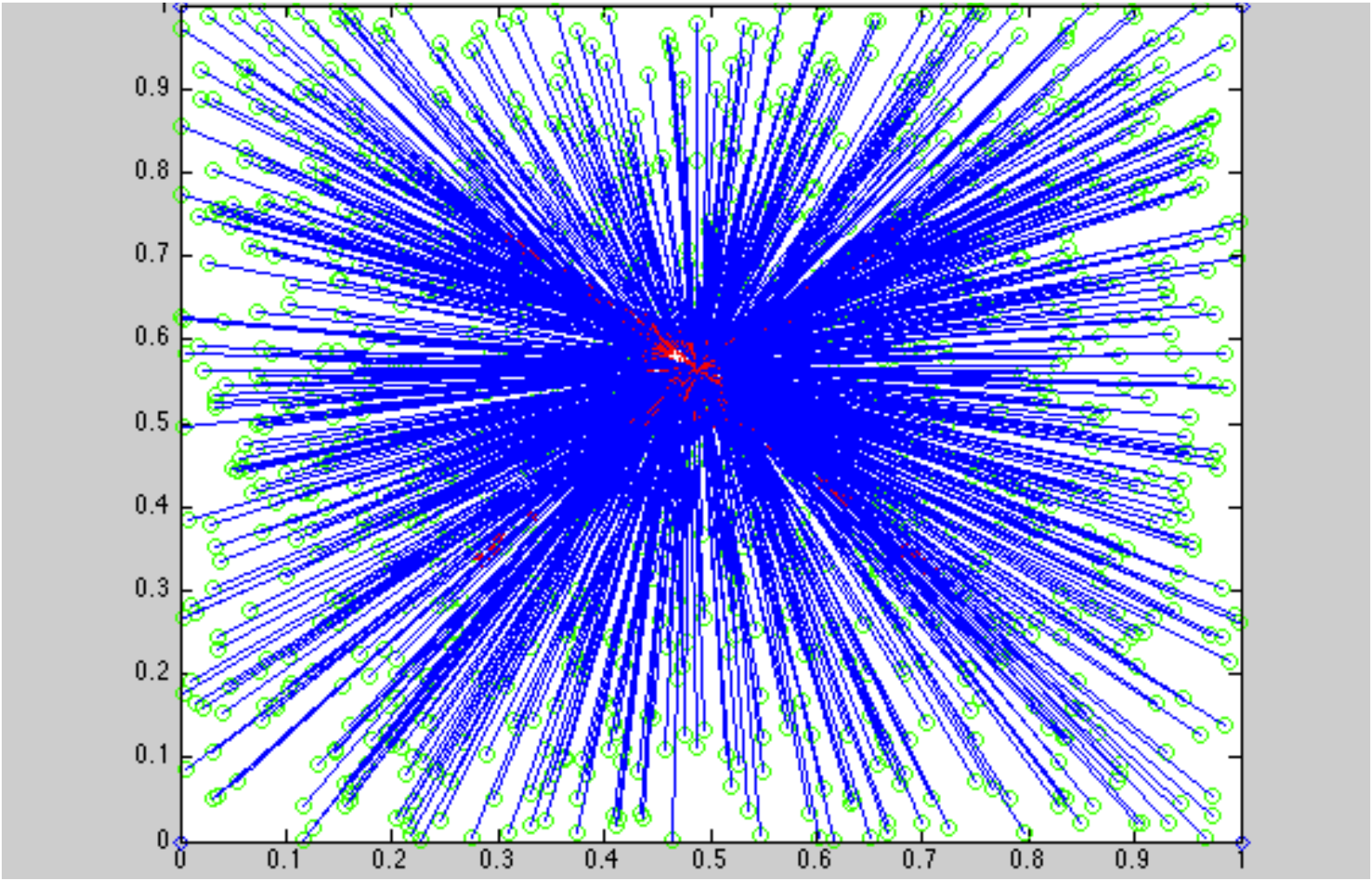
$$\begin{pmatrix} \nabla_x \Xi(x,\sigma) \\ -\nabla_\sigma \Xi(x,\sigma) \end{pmatrix} = 0, \sigma \geq 0$$

Numerical Results

- The experiments has been carried out on 2-D and 3-D networks
- The number of sensors ranges from 1000 to 15000 for 2-D problems and from 1000 to 5000 for 3-D problems
- The software used for benchmark is SFSDP
  - A formulation based on the SD relaxation of the Sensor Network Localization problem
  - That reduces the number of connections and consequently the calculations in the algorithm
  - The algorithm used to solve this relaxation is SeDuMi
- In order to understand the goodness of the solution found we use the Square Root Distance

$$RMSD = \left(\frac{1}{N} \sum_{p=1}^{N} \left\| x_p - x_p^* \right\|^2 \right)^{\frac{1}{2}}$$

Numerical Results



Numerical Results

The anchors are located in the corners:

		IPM		SFSDP		
N	m	RMDS	Time	m Red.	RMDS	Time
1000	11903	7.89E-07	1.55	3669	2.08E-04	37
3000	47848	2.23E-10	10	11669	1.37E-04	95
5000	83848	2.95E-10	24	19669	1.16E-04	157
10000	173848	6.43E-10	102	39669	5.00E-05	341
15000	263848	2.81E-12	255	Out of Memory		

- High precision
- Very small elapsed time
- Capable of getting good results even on high dimensional problems

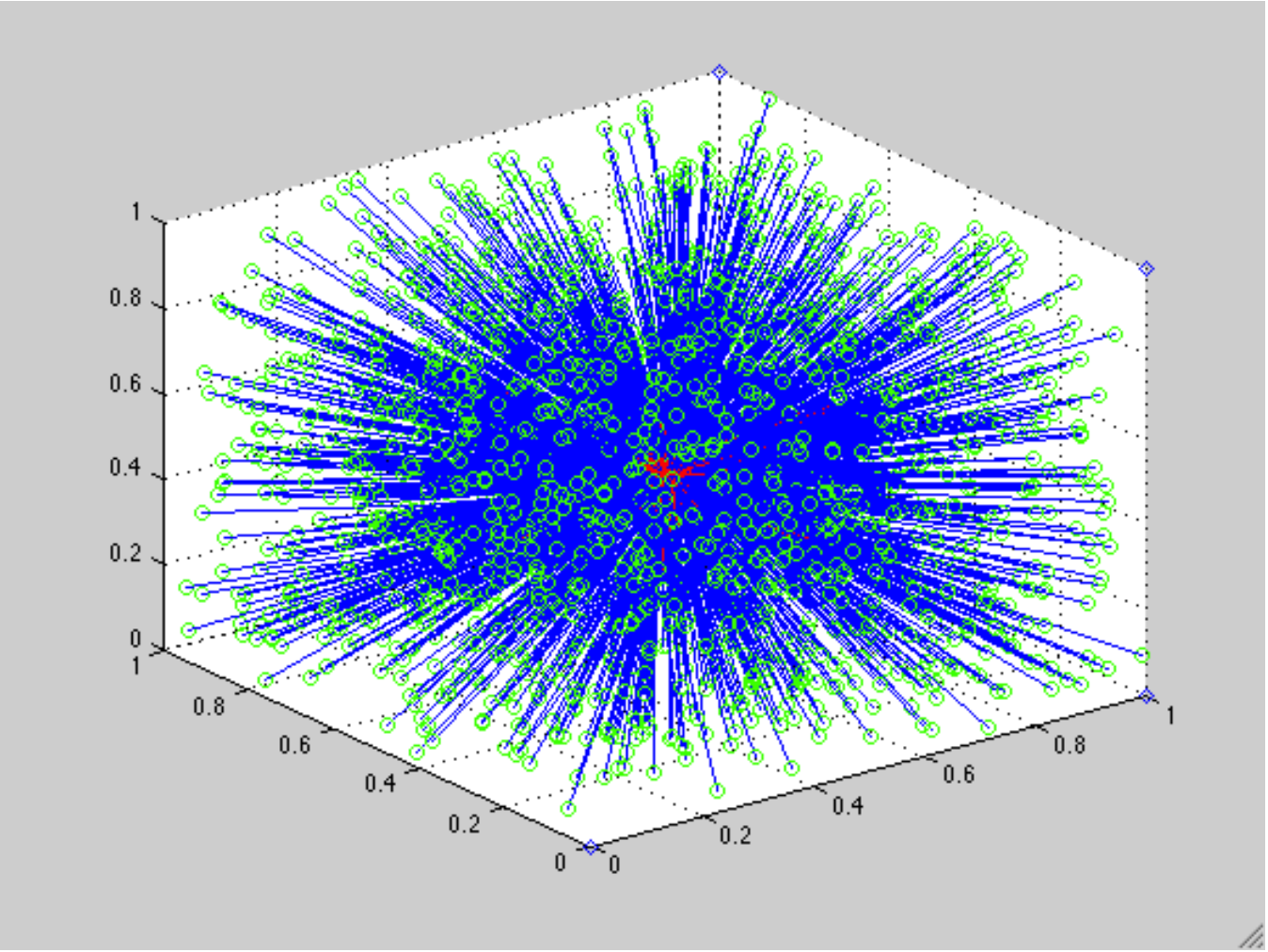
Numerical Results

With 50 anchors randomly distributed:

		IPM		SFSDP		
N	m	RMDS	Time	m Red.	RMDS	Time
1000	12671	3.23E-06	3	4022	6.24E-05	18
3000	48659	3.31E-07	15	12290	6.92E-05	60
5000	84599	0.0453	114	20879	7.49E-05	156
10000	174668	4.68E-06	188	41626	6.24E-05	243
15000	265668	1.46E-07	500	Out of Memory		

- The problem is quite more difficult from a numerical point of view
- Without anchors in the corners there is a translational degree of freedom
- SFSDP reduces the translational freedom

Numerical Results 3-D



Numerical Results 3-D

		IPM		SFSDP		
N	m	RMDS	Time	m Red.	RMDS	Time
1000	17296	2.15E-09	4	4700	3.76E-04	216
2000	41272	4.40E-11	11	9700	3.45E-05	402
3000	65263	3.20E-12	20	14700	6.14E-06	608
4000	89263	1.30E-12	32	19700	7.86E-06	925
5000	113263	6.59E-13	47	24700	2.41E-05	1333
1000	19004	5.43E-06	6	2600	1.09E-05	18
2000	43048	4.99E-06	14	5100	1.70E-05	33
3000	66983	5.20E-12	26	16900	1.41E-05	67
4000	91046	0.0012	106	22000	4.76E-06	91
5000	114993	0.0123	188	28000	7.73E-06	127

- Same behaviour of the 2-D problem

Noise

- In many sensors network localization problems the measurements on the distances are affected by noise
- The presence of noise greatly complicates the problem
- We add noise to the instances in order to test the robustness of the algorithm
- The Noise is Gaussian with variance equal to 0.2

Noise 2-D

		IPM		SFSDP		
N	m	RMDS	Time	m Red.	RMDS	Time
1000	13479	6.40E-03	9	3700	1.89E-02	46
3000	61161	6.90E-03	59	11600	6.38E-03	217
5000	109161	6.90E-03	116	19700	8.28E-03	425
10000	229161	7.30E-03	440	39650	6.50E-03	1391
15000	349161	7.30E-03	1293	Out of Memory		
1000	14577	8.70E-03	6	4000	1.82E-02	39
3000	62311	5.60E-03	26	12200	5.81E-03	140
5000	110283	1.30E-02	75	20900	6.91E-02	505
10000	230304	5.50E-03	252	41000	8.25E-03	1128
15000	350314	1.63E-02	637	Out of Memory		

- The Noise significantly increase the difficulty of the problem
- The IPM by getting near the global minimum yields better results

Noise 3-D

		IPM		SFSDP		
N	m	RMDS	Time	m Red.	RMDS	Time
1000	20427	6.40E-03	15	4695	3.90E-02	257
2000	52249	6.90E-03	45	9589	3.11E-02	647
3000	84220	6.90E-03	88	14522	2.50E-02	1287
4000	116220	7.30E-03	131	19700	2.23E-02	1831
5000	148220	7.30E-03	225	24700	2.37E-02	2819
1000	23308	8.70E-03	12	5600	3.12E-02	43
2000	55121	5.60E-03	33	11000	2.94E-02	157
3000	87185	1.30E-02	60	16900	2.28E-02	145
4000	119119	5.50E-03	97	22000	1.99E-02	277
5000	151097	1.63E-02	145	28000	3.79E-02	581

- Results better than the 2-D problem
- The fact that the result is more accurate is really important because the anchorless version of this problem corresponds to the geometry predictions of molecules

Conclusions

- The results show that the proposed method can be very useful to solve large instances of problems from Canonical Duality
- For the sensor localization problem the algorithm compares well even if written in Matlab, an interpreted language
- In the future the newfound method can:
  - Be used to efficiently find solutions for problems formulated with canonical duality theory
  - Be improved in order to become the best method to solve the sensor network localization problem
  - It is also possible to write the code of a method that:
    - Erases the translational degree of freedom
    - Works for the anchorless problem

BIBLIOGRAPHY

- http://uob-community.ballarat.edu.au/~dgao/papers/paper.html
- http://uob-community.ballarat.edu.au/~dgao/papers/CACE-09.pdf
- Gao, D.Y., *Duality Principles in Nonconvex Systems: Theory, Methods and Applications*. Kluwer Academic Publishers, Boston/Dordrecht/London, 2000, xviii+454pp.
- V. Latorre: A Potential Reduction Method for Canonical Duality, with