APPLIED SPARSE OPTIMIZATION TECHNIQUES

School of Doctorate in Industrial Technologies, Research in Mathematical Optimization,

Santiago González Jorge,

(1) Introduction: *Sparse Optimization, Why are we interested?*

Sparse optimization is related to the task of analysing and processing databases where the amount of data and the amount of valid information differ largely. It has been used to represent, compress and estimate both static and dynamic data. Typical scenarios include problems where data available is insufficient or, on the contrary, overwhelming, and still it is required to retrieve meaningful information. *Big-data analysis, data-mining, compressed sensing, sparse-coding pose problems related to solution spaces with a large number of dimensions, often subject to non-linear sets of restrictions and non-linear target functions.*

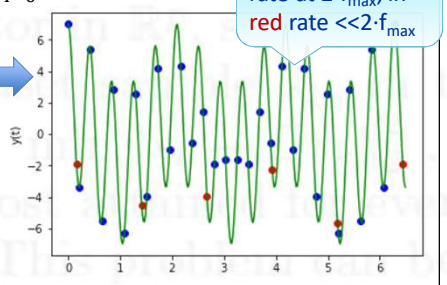
(2) The problem: *Signal reconstruction from samples, classic approach fails at subsampling*

Nyquist–Shannon: *Sufficient condition to reproduce a continuous-time signal → sampling at rate f_s higher than 2 times its f_{max}. Then, use Fourier approximation*

$$y'(t) \approx \sum_{i=0}^{n} x_i \cos(2\pi f_0 i \cdot t)$$

in blue sampling rate at $2 \cdot f_{max}$; in red rate $<< 2 \cdot f_{max}$

…But, ¿Can we reconstruct the signal when the sampling rate f_s is much lower than $2 \cdot f_{max}$?



(3) The solution: *Signal reconstruction by sparse coding approach. Convex Problem*

Subsampling f_s on our example would only allow a Fourier approximation up to the 4^{th} harmonic, which fails four our signal with a 15^{th} harmonic:

$$y(t) = 2 \cdot cos(3t) + 5 \cdot cos(15t) \quad \text{…here } 2\pi f_0 = 1$$

Sparse coding, *in the other hand, propose an overcomplete basis (**dictionary**) whose elements (**atoms**) are the vectors c^i, with components $cos(i \cdot t_s)$, taken at the sampling instants $\{t_s\}$. $y(t)$ may be approximated by $\sum_{i=0}^{n} c^i \cdot x_i$ (i.e. $y \approx Cx$) at the sampling spots and by $\sum_{i=0}^{n} x_i \cdot cos(it)$ elsewhere*

In our example n could be 20

…With more harmonics (n) than samples at $\{t_s\}$, $y = Cx$ is underdetermined and L.S.M. leads to many c^i being present at the solution (its $x_i \neq 0$), to find out one with only a few, we solve a sparse optimization problem:

(SP): $min \|x\|_0$ *s.t.* $\|y - Cx\|_2 <$

L_0 pseudo-norm of x (i.e. $\|x\|_0$) counts the number of non-zero components of x, therefore the number of atoms being present at the solution

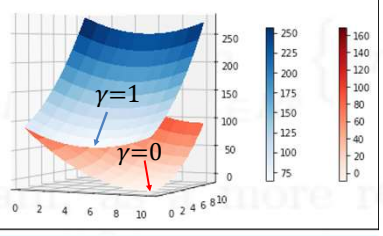
…But the above problem, due to $\|x\|_0$, is non-linear and non-convex, in general, and it is often approximated with the help of the L_1 norm by:

(SP'): $min \|y - Cx\|_2 + \gamma \cdot \|x\|_1$ …with $\gamma > 0$

(4) The downside: *Solution is dependent on the coefficients used to build up the target function*

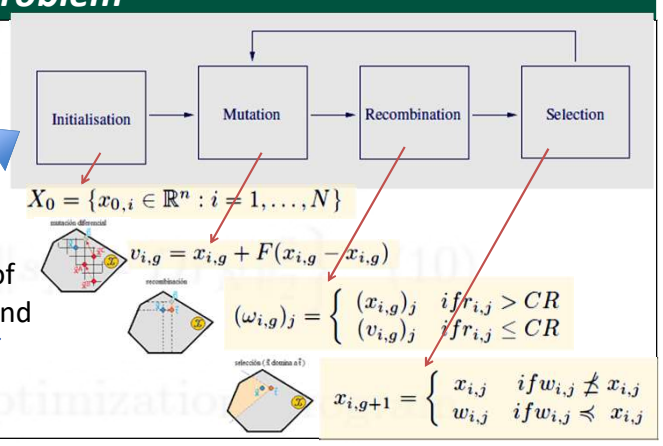
Approximated problem (SP') shown in (3) is convex when γ is properly chosen, generally leading to the sparse solution…

The problem: The solution is very sensitive to γ; it may exhibit a large quadratic error $\|y - Cx\|_2$, it may even not exist.



(5) A possible overcome : *Multi-step Multi-objective Differential Evolution. Non-Convex Problem*

- Ideally we would like to obtain the set of solutions of the original problem (SP): $min \|x\|_0$ *s.t.* $\|y - C\|_2 < \mathcal{E}$
- But solving (SP) by exhaustive search is extremely inefficient… *although there are certain type of evolutionary algorithms, namely **differential evolution (DE)**, has been effectively used in similar scenarios.*
- DE convergence depend on the number of components of the solution, population size and algorithm parameters and policies *…Multi-step approach first reduce the number of active components on the solution and then tries to improve the precision of the result iteratively.*



(6) An application example: *Energy Disaggregation*

Energy Disaggregation via Discriminative Sparse Coding is a promising application of sparse optimization.

In the original setting proposed by Kolter-Siddarth, during the training phase, the energy consumption of a certain device is measured hourly, during a week, in a set of M different households, giving the matrix X_i:

$$X_i = \begin{bmatrix} x_i^{1,1} & x_i^{1,2} & \cdots & x_i^{1,M} \\ x_i^{2,1} & x_i^{2,2} & \cdots & x_i^{2,M} \\ \vdots & \vdots & \ddots & \vdots \\ x_i^{T,1} & x_i^{T,2} & \cdots & x_i^{T,M} \end{bmatrix} = \left[x_i^{(1)}, \ldots, x_i^{(M)}\right]$$

… this is repeated for all the typical home appliances.

(7) Energy disaggregation: *The training phase*

During the training phase we have gathered the matrixes X_i of energy consumption for each kind of appliance.

… In the simplest approach, we use this information to set an sparse optimization problem. Its target is to find out the best basis functions B_i to encode de consumption X_i related to each appliance.

$$min \ \|X_i - B_i A_i\|_F^2 + \sum_{n,m} A_i[n,m]$$
$$s.t. \quad \|B_i\|_2 \leq 1$$

Frobenius norm $\| \ \|_F$ generalizes norm-L_2 on a matrix. The second term generalizes L_1

…The problem above is not simultaneously convex on A_i and B_i, but it is convex in each one if we keep the other variable constant.

(8) Energy disaggregation: *The Test phase*

At the test phase we are finally presented to a real aggregated energy consumption X adding up all the appliances for each household. Our final interest is to retrieve the approximate disaggregated consumptions X_i for each kind of appliance.

… In the simplest approach, we set an sparse opt. problem to find out the approximate coefficients A_i:

$$\begin{bmatrix} \hat{A}_1 \\ \vdots \\ \hat{A}_K \end{bmatrix} = \arg \min \left\| \overline{X} - [B_1, \ldots, B_K] \cdot \begin{bmatrix} A_1 \\ \vdots \\ A_K \end{bmatrix} \right\|_F^2 + \lambda \sum_{i=1}^{K} \|A_i\|_1$$

…Then we can estimate the energy consumption of each appliance:

$$\hat{X}_i = B_i \cdot \hat{A}_i$$

(9) Summary: *Sparse Optimization, Energy disaggregation and next Milestones*

- **Energy disaggregation** provides a good example for the potential of **sparse optimization algorithms**, nevertheless, even by resorting to more sophisticated approaches than the one outlined here, the algorithms achieve accuracies around the 50%. This performance is related to low time resolution of measurements and is improved by increasing the sampling rate, as shown in applications of hidden additive Markov models, with accuracies close to 90%. …*Next milestones consist on developing more sophisticated models, including unsupervised learning and replacing the L_1 norm by the original L_0.*
- Sparse optimization is suited for a huge set of applications such as data recovery, image & speech recognition, data compression, big-data analysis, etc. *Among them, **inverse problem** resolution might be one of its most promising applications, due to its singular characteristics.*
- Current state of the art is related to the use of the L_1 norm and the resolution of convex problems. *In these scenarios multi-step multiobjective differential evolution might help to recover the original formulation based on the L_0 norm and confronting non-convex & non-linear problems.*