Top (2009) 17: 215–236
DOI 10.1007/s11750-009-0124-1

INVITED PAPER

Optimization and data mining in medicine

Panos M. Pardalos · Vera Tomaino · Petros Xanthopoulos

Published online: 8 December 2009

Abstract Mathematical theory of optimization has found many applications in the area of medicine over the last few decades. Several data analysis and decision making problems in medicine can be formulated using optimization and data mining techniques. The significance of the mathematical models is greatly realized in the recent years owing to the growing technological capabilities and the large amounts of data available. In this paper, we attempt to give a brief overview of some of the most interesting applications of mathematical programming and data mining in medicine. In the overview, we include applications like radiation therapy treatment, microarray data analysis, and computational neuroscience.

Keywords Optimization · Medicine · Computational biology · Computational neuroscience

Mathematics Subject Classification (2000) 90-02 · 90C90 · 92B05

This invited paper is discussed in the comments available at: doi:10.1007/s11750-009-0125-0, doi:10.1007/s11750-009-0127-y, doi:10.1007/s11750-009-0128-x, doi:10.1007/s11750-009-0129-9, doi:10.1007/s11750-009-0130-3.

P.M. Pardalos (✉) · V. Tomaino · P. Xanthopoulos
Industrial and Systems Engineering, University of Florida, Gainesville, FL 32601, USA
e-mail: pardalos@ufl.edu

P. Xanthopoulos
e-mail: petrosx@ufl.edu

V. Tomaino
Department of Experimental and Clinical Medicine, University Magna Græcia of Catanzaro, Catanzaro, Italy
e-mail: vera.tomaino@unicz.it

1 Introduction

Mathematical programming and optimization were initially employed in order to address many problems related to personnel scheduling, supply chain management, manufacturing process design and other applications of high industrial interest. Lately researchers have discovered new problems in areas that no one could imagine a few decades ago. One of these areas is medicine. In recent years, more and more statisticians, operations researchers and data miners have tried to apply their knowledge to problems posed by the medical community. This has led to the development of a blooming area that lies in the intersection of mathematics and medicine. Furthermore, the amount of literature in this area has increased dramatically over the last few years (Pardalos and Romeijn 2009; Seref et al. 2008; Mondaini and Pardalos 2008; Alves et al. 2008; Lim and Lee 2008; Pardalos et al. 2004b, 2007; Pardalos and Principe 2002; Floudas and Pardalos 2000; Pardalos and Hansen 2008).

The formation of this new cross-disciplinary area of research was accelerated mainly by two important factors: (i) the recent biological advances which produced a vast amount of genome data and (ii) the technological development of the medical instrumentation which has made the interface between the two areas possible (e.g., modern medical devices come combined with embedded programming language libraries that make it possible to apply the results of theoretical modeling in a real clinical setup).

Both medicine and optimization define two very broad fields of research. However, medical applications are not limited just to clinical practice, but they also include drug design, study of complex parts (e.g., human brain), interpretation of genome expression data, etc. On the other side, the general optimization problem

$$\min_{x \in X} f(x)$$

refers to very large categories of "easy" and "difficult" problems depending on the structure of the objective function $f(x)$ and the properties of the feasible set X. In this paper, we attempt to present some of the most interesting applications of optimization and data mining in the field of medicine.

This area is very broad and is very difficult to present all the research directions developed over the last few years in a survey paper. However, we try to give a taste of this multidisciplinary field especially to researchers who are interested to apply mathematical tools in new problems. Hence, we have chosen to present four relevant problems together with the main algorithmic approaches and their open challenges.

The structure of the rest of the paper is as follows: In Sect. 2, we present the main optimization challenges in the area of radiation treatment planning together with the most prominent optimization frameworks. In Sect. 3, we discuss the application of biclustering in the data mining related to the problem of microarray data processing. In Sect. 4, we discuss two applications taken from the field of computational neuroscience: an application of multiquadratic 0–1 programming as part of an epileptic seizure prediction algorithm and the general framework of the electroencephalographic inverse problems which are studied in the field of computational neuroscience. We conclude with some remarks on the challenges and the future directions of this cross disciplinary area of research.

2 Radiation therapy

Radiation therapy (RT) is one of the most popular treatments for cancer. Radiotherapy, depending on the stage, location and the type of tumor, can be used as palliative or curative intention, alone or in combination with the other treatment plans such as surgery and chemotherapy. RT consists of transmitting high energy beams directly to the patient's tumor location. RT is delivered to the patients through radiation beams that are produced by a linear accelerator with Multileaf Collimators (MLCs) (Kamath et al. 2009; Kalinowski 2008; Lim et al. 2007) placed on the head of the treatment unit. MLCs can give a shape to the radiation beams by moving their leaves through a computer controlled interface, so that they match the shape of the tumor. The view that the beam source can see from the MLCs is called *Beam's-Eye View* (BEV) of the target (Aleman et al. 2008; Lim et al. 2007). BEV ensures more irradiations at the tumor and less at the healthy tissue.

The major problem of this treatment is that it destroys not only the tumor cells, but also the surrounding healthy tissue. As a consequence, the main challenge of the radiation therapy is to send an appropriate dose into the *Planning Target Volume(s)* (PTVs) (i.e., the tumor and the possibly involved tissue), and in the meantime to spare the *Organs At Risk* (OARs). OARs are the organs that are not affected by the tumor, but close to it, and that should be protected from the radiation in order to avoid medical complications. Hereby, it is very important to identify the three-dimensional (3D) shapes of organs that are going to receive the treatment. In optimization modeling, the 3D volume is represented by a grid of voxels. Another important parameter that the radiation therapy practitioner has to take into consideration is that the volume and the shape of the tumor can change during the treatment, and also the organs close to the tumor may change their shape due to the possible patient's loss of weight throughout the treatment process. Furthermore, there exists another source of uncertainty which is caused by the patient motion during treatment (for example, due to breathing motion) and it is called *intrafraction motion* (Romeijn and Dempsey 2008).

Using 3D imaging techniques, such as Computer Tomography (CT) or Magnetic Resonance Imaging (MRI), it is possible to have a clear image of the location and the volume of the tumor. The 3D images allow the physician to identify the Gross Tumor Volume (GTV), the Clinical Target Volume (CTV) (meaning the volume of possible spread), the PTV and the OARs. Therefore, manual treatment planning is extremely tedious because the clinician has to determine several treatment parameters manually. Furthermore, it is important for a treatment plan to have uniform dose distribution on the target to minimize the formation of *hot* and *cold spots*. *Hot spots* are the tissues that receive more than the prescribed dose whereas *cold spots* are the tissues that receive less than the prescribed dose. Hence, radiation therapy treatment planning is a challenging optimization application because one has to embody all the aspects of the problem and model them as decision variables and constraints.

There exist two types of radiation treatment planning: *forward planning* and *inverse planning*. The *forward planning* is a trial-and-error approach where, if the result is not adequate, the plan parameters are manually changed and the process starts over. This is, in general, a very long process and it does not always achieve high quality

treatment effects. On the other hand, in the *inverse planning*, an objective function measures the quality of the treatment plan. Nowadays more people have started to use the inverse treatment planning (also called *computer-based treatment planning*). Treatment goals can be modeled using a variety of optimization techniques: linear programming, nonlinear programming, mixed integer programming, and dynamic programming.

Although there exist many types of radiation therapy as Tomotherapy, Proton Therapy, Intensity Modulated Proton Therapy and so forth, in this paper we are going to give a brief overview of two of the most promising methods: *Intensity Modulated Radiation Therapy* (IMRT) and *Three-Dimensional Conformal Radiation Therapy* (3DCRT). In spite of a wide spectrum of proposed approaches for solving these problems, in this paper, we focus on the state-of-the-art methods with respect to the computational results and treatment effectiveness.

2.1 Intensity modulated radio therapy treatment planning

IMRT modeling is the most time consuming treatment planning (with respect to computational time needed), but at the same time it is the one that delivers best quality of treatment. The inverse treatment planning problem of IMRT allows several beams with different shapes and different intensities to reach the target region. In fact, beam intensity as well as dose conformity may vary simultaneously (Reemtsen and Albert 2009; Lim 2008). In 3DCRT, the intensity of the radiation during the beam's cross-transmission is uniform, which can be modified by the use of pre-fabricated wedge filter. However, this technique (especially for targets that have irregular shapes and that are non-convex, like para-spinal tumors) is limiting with respect to the delivery of the right dose distribution at the tissue that fits the shape of the tumor (Küfer et al. 2009). IMRT has the potential to deliver a more accurate dose distribution to the target volume, while sparing the normal tissue, more than the other treatment techniques, in particular for target volumes and/or organs at risk with complex shapes and/or concave regions (Intensity Modulated Radiation Therapy Collaborative Working Group 2001).

The two most common techniques of IMRT using the MLCs are: Segmental Multileaf Collimator (SMLC) and Dynamic Multileaf Collimator (DMLC) (Kamath et al. 2009). In SMLC, the beam is switched off while the leaves are in motion, whereas in DMLC, the beam is switched on at the beginning of the treatment and switched off at the end of the treatment. In IMRT, Beam Orientation Optimization (BOO) is the problem to decide which beam orientation should irradiate the patient. Aleman et al. (2008) explained in detail how to solve the BOO problem. The goal of IMRT is to find, for each beamlet and according to the optimization goals of the respective model, a suitable nonnegative beamlet weight defining its radiation intensity (Reemtsen and Albert 2009).

One of the most popular subproblems of IMRT is the *Fluence Map Optimization* problem (Aleman et al. 2007, 2008) which has to quantify the quality of the treatment plan and find radiation intensities for all beamlets. Let B represent the set of the candidate beam orientations from which radiation may be delivered and $\theta \in B^k$ a predetermined set of beam solutions where k is the number of beams in θ and

$B^k = B \times B \times \cdots \times B$ (k times). B_θ denotes the set of all bixels (beamlet intensities) in θ. S denotes the total number of structures where the targets are indexed as $s = 1, \ldots, T$ and the critical organs are $s = T + 1, \ldots, S$. Furthermore, each structure s is discretized into a finite number v_s of voxels. The dose received by voxel j in structure s is denoted by z_{js}. It can be computed according to the following formula:

$$z_{js} = \sum_{h=1}^{k} \sum_{i \in B_{\theta_h}} D_{ijs} x_i, \quad j = 1, \ldots, v_s, \ s = 1, \ldots, S,$$

where D_{ijs} represents the dose absorbed by the volume element j (voxel) in structure s from beamlet i at unit intensity and x_i represents the intensity of bixel i. Therefore, a basic formulation of the FMO problem is given by:

$$\begin{aligned} \min \quad & \sum_{s=1}^{S} \sum_{j=1}^{v_s} F_{js}(z_{js}) \\ \text{s.t.} \quad & z_{js} = \sum_{h=1}^{k} \sum_{i \in B_{\theta_h}} D_{ijs} x_i, \quad j = 1, \ldots, v_s, s = 1, \ldots, S, \\ & x_i \geq 0, \quad i \in B_{\theta_h}, h = 1, \ldots, k. \end{aligned} \tag{1}$$

The objective function $F_{js}(\cdot)$ denotes a convex penalty function for voxel j in structure s. Romeijn and Dempsey (2008) propose an interesting approach that associates radiation treatment planning and risk management, meaning that both problems try to minimize a loss function. While in the risk management the loss function is defined as a perceived risk, in radiation therapy it can be defined as a deviation from the prescribed radiation dose for each organ.

2.2 The three-dimensional conformal radiation therapy

The 3DCRT also uses a MLC but in this case the *leaves* are fixed during the treatment, shaping the beam of radiation for each angle so that the tumor will be hit with respect to its exact shape. Lim (2009) explains several advantages of this approach compared to IMRT. This latter approach is able to give a high quality treatment plan, but it is more difficult than 3DCRT. This is because the IMRT-based optimization model considers all the beamlets. Also for this reason, the computational time to obtain the solution is higher. Furthermore, the IMRT can easily give a discretization error to the model due to the several beams used for the treatment.

Usually, a single beam is used for superficial cancer, whereas for deeper cancer cross-firing beams are used delivered from different directions. The goal in conformal therapy is to achieve high probability of tumor control while minimizing the damage to the tissues close to the tumor. Especially for the cancer close to a curved surface, the wedge filter, a fine metallic block with a thick side (the heel) and the thin edge (the toe), is particularly useful (Lim 2007, 2008). When the edge is placed in front of the aperture, less radiation is transmitted through the heel than through the toe. It

is possible, regarding the treatment of the patient, to use different wedges or to use a permanent wedge located at the head of the treatment unit or to rotate the wedge to the desiderated orientation or remove it completely.

The quality of a treatment plan is usually evaluated through a dose–volume histogram (DVH) (Lim 2009; Küfer et al. 2009), a plot which shows what fraction of volume receives dosage in some range. In other words, this histogram specifies the fraction of the volume of a structure that receives more than a certain amount of dose (Romeijn and Dempsey 2008). Next, DVH is used to find quality solution in clinical perspective, while in this paper we discuss some techniques for reducing computational time.

2.2.1 Optimization models for 3DCRT

In these optimization models, $D_{(i,j,k),A}$ is used to indicate the dose contribution to voxel (i, j, k) from a beam of unitary weight from angle A. We use the following notations:

- A is the set of the beam angles
- T is the set of the voxels that comprise the PTV
- S is the set of voxels in the OARs
- N is the set of voxels in the normal tissue
- ω_A is the beam weight delivered from the angle A

Furthermore, λ_t, λ_s and λ_n represent the relative weighting factors for the PTV, OARs and normal tissue, correspondingly, in the objective function. If the wedge filter is used, the dose can be modeled by $D_{(i,j,k),A,F}$ which represents the dose contribution to voxel (i, j, k) from a beam of unitary weight from angle A using wedge orientation F.

Optimization of the beam weights The classical optimization problem in conformal radiation therapy is to choose which weights are delivered from a given set of angles. The total dose $D_{(i,j,k)}$ to voxels (i, j, k) is obtained by the sum of the contributions of all the angles $A \in \mathsf{A}$. Given a set $\Omega = T \cup S \cup N$, a general optimization model that determines optimal radiation intensity is:

$$\begin{aligned} \min_{\omega} \quad & \lambda_t f(D_T) + \lambda_s f(D_S) + \lambda_n f(D_N) \\ \text{s.t.} \quad & D_\Omega = \sum_{A \in \mathsf{A}} D_{\Omega,A}\omega_A, \\ & l \le D_T \le u, \\ & \omega_A \ge 0, \quad \forall A \in \mathsf{A}. \end{aligned} \tag{2}$$

The upper and lower bound constraints are imposed on the target dose in such a way that, in the worst case, the solution will satisfy the minimum requirement for a treatment plan. Usually, the objective function $f(D)$ can be chosen with respect to the treatment plan but the general function can be defined as:

$$f(D) = \left\| D(\cdot) - \theta \right\|_p, \quad p \in \{1, 2, \infty\}. \tag{3}$$

Here θ is the prescribed dose for an organ's treatment.

Optimization beam angle and wedge orientation Lim (2008) proposes an optimization model that optimizes beam angles, wedge orientations and beam intensities. As mentioned above, the orientation of the wedge is very helpful to reduce the radiation dose at the OARs. To consider that, it adds an extra dimension F to the variable ω_A. Let us consider K as the maximum number of beam angles. The Mixed Integer Programming (MIP) formulation is as follows:

$$\begin{aligned}
\min_{\omega,\psi} \quad & \lambda_t f(D_T) + \lambda_s f(D_S) + \lambda_n f(D_N) \\
\text{s.t.} \quad & D_\Omega = \sum_{A \in \mathsf{A}} D_{\Omega,A,F}\omega_{A,F}, \\
& \omega_{A,F} \leq M \cdot \psi_A, \\
& l \leq D_T \leq u, \\
& \sum_{a \in \mathsf{A}} \psi_a \leq K, \\
& \psi_A \in \{0,1\}, \quad \forall A \in \mathsf{A}.
\end{aligned} \tag{4}$$

The binary variables ψ_A indicate whether or not an angle A is used for the treatment. The constraint $\omega_{A,F} \leq M \cdot \psi_A$ (where M is a sufficiently large constant) ensures that the weight of $\omega_{A,F}$ is nonzero only if $\psi_A = 1$. It is important to choose a good value of M because it is crucial for the computational speed of the model. In most cases, the goal is to find a minimum number of beams that satisfy the treatment plan. In general, a treatment plan that involves few beams (i.e., 3 to 5) is preferred over the one that involves several beams because in this case, less time is required for the treatment. Consequently, it is important to consider that if many beams are used (more than 5) then the beam orientation becomes less important for the optimization model.

2.2.2 Solution approaches

It is possible to reduce the solution time or decrease the problem size considering fewer voxels (because usually not all of them receive radiation) or reduce the beam angles. For the first case, it is possible to use a normal tissue voxel reduction approach (preprocessing and reducing resolution in the normal tissue) and, for the latter one, an alternative approach is proposed by Lim et al. (2007) which is called *three-phase approach*. This is a multiphase approach similar to the one that Legras et al. (1982) have proposed, which achieves the solution after solving a sequence of models. The models are solved in ascending order of difficulty and the solution of one model is the starting point of the next one. Each model is different from the rest, with respect to the number of the selected voxels, including the formulation and the number of beams. In this approach, the most promising beam angles are identified so that the discrete variables are reduced in number (i.e., the beam angles that pass directly through any

OAR are removed). Another more elaborated approach considers a score function for every candidate angle that is able to deliver a high dose to the PTV, and avoid or deliver a lower dose to the OAR or to the normal tissues. For this reason, only the beam angles with a high score are included in the model (Pugachev and Xing 2001). Below, the three-phase approach, as incremental modeling scheme, is given (Lim et al. 2007; Lim 2008, 2009). This approach achieves a near-optimal solution, and it needs less time to solve the problem than the original formulation.

- *Phase 1. Selection of promising beam angles*. The goal in this step is to choose a subset of beam angles A_1, a set of all angles $A \in \mathsf{A}$ for which $\omega_A > 0$ for at least one of these r sampled problems. This can be achieved by solving a set of r MIPs, where each of them is formed using only a reduced collection of voxels from the ones in: the PTV, a random selection of voxels including 10% of the OAR (S') and the voxels in $R_\rho(T)$ (Lim et al. 2007; Lim 2008, 2009). So it is possible to define Ω_1 as

$$\Omega_1 = \left\{T \cup S' \cup R_\rho(T)\right\},$$

  where $R_\rho(T)$, for a parameter ρ, denotes a subset of normal region voxels close to the PTV. In other words,

$$R_\rho(T) := \left\{(i, j, k) \in N | \text{dist}\big((i, j, k), T\big) \leq \rho\right\},$$

  where $\text{dist}((i, j, k), T)$ is the Euclidean distance from the center of the voxels (i, j, k) to the PTV.
- *Phase 2. Treatment beam angle determination*. It solves Model (4) considering only the voxels that are close to the PTV or that stay in the OAR or in the reduced normal region, described by the following:

$$\Omega_2 = \left\{T \cup S \cup R_\rho(T) \cup N_1\right\},$$

  where N_1 is a reduced region of the normal one and its voxels (i, j, k) are only even. This means

$$N_1 := \left\{(i, j, k) \in N | i \bmod 2 = j \bmod 2 = k \bmod 2 = 0\right\}. \tag{5}$$

- *Phase 3. Final approximation*. It solves the reduced optimization problem over the complete set of voxels considering K beam angles. For that, it fixes $\psi_{A_1} = 1$ for the angles selected in phase 2 and $\psi_A = 0$ otherwise. This approximation is easier to solve compared to the full optimization model because of both data reduction and binary variable elimination purposes.

Lim et al. (2007) highlight, based on computational experiments, that the quality of the approximation of this approach is close to the optimal. However, there is no guarantee that this technique will give the same result as the full Model (4).

3 Microarray data analysis

In life sciences, the importance of data analysis rapidly emerges due to the increasing volume of data that becomes available every day (e.g., Human Genome Project, etc.).

The discovery of novel mathematical tools has made it possible to analyze these large datasets to some extent, but still there are many challenging open problems.

One biological tool that boosted the generation of gene expression data is the *DNA microarray technology*. These devices allow researchers to observe how genes act in different types of cells and under various conditions. As a matter of fact, now it is possible to analyze not only how a gene behaves itself but also how it interacts with other genes and which gene expression patterns are formed under predefined conditions. Experiments with microarrays can identify and classify genes responsible of certain medical states, but they are especially crucial for understanding *genetic diseases* caused by mutations, in a gene or a set of genes, that make the mutant genes inappropriately expressed or even not expressed at all. For example, it is known that several cancer types can be caused by inactivation, deletion or, on the contrary, by constitutive activity of *p53 tumor suppressor gene* (Busygin and Pardalos 2007).

The analysis of the microarray data is not an easy task. The most important factors that impede the success of the microarray technology are: the high dimensionality of the data, poor accuracy of microarray probes and the relevant difficulties to sample probes (i.e., the procedure might be very painful for alive patients, while the gene expression level rapidly degrades in dead tissue).

It is possible to divide the data mining problems into four major classes that sometimes overlap: statistical analysis, clustering/classification, dimensionality reduction and biclustering (or co-clustering). Statistical analysis for microarray data usually consists in calculating the *fold change* (the ratio of the measured value for an experimental sample to the value for the control sample) of particular genes across different groups of samples applying classical statistical tests such as t-test, ANOVA, Wilcoxon test, etc. These techniques are appropriate when a proper separation of samples into classes is known, the number of outliers in each class is insignificant and the data are assumed to have certain statistical properties (e.g., normal distribution).

Supervised learning techniques, also called *classification*, require a complete labeled *training set* from which the classification algorithm learns or it is trained to classify unknown samples. Examples of supervised learning techniques are: linear discriminant analysis, Support Vector Machines (SVMs), regression tree, etc.

On the other hand, unsupervised techniques (also called *clustering*) do not require a training set but rather identify common sample properties and cluster the data based on them. Examples of clustering techniques are: k-means clustering, hierarchical clustering, self-organizing maps (SOM), spectral graph clustering, and partitioning clustering. Choosing "the best" algorithm for a particular problem may not be always a trivial task. There are multiple clustering techniques that can be used to analyze expression data. Advantages and limitations may depend on factors such as the statistical nature of the data, pre-processing procedures, number of features, etc. Moreover, it is common to observe inconsistent results when different clustering methods are tested on a particular data set. Usually, to avoid the computational time and memory problems, partitioning clustering is preferred to hierarchical clustering, especially in large-scale problems. If the number of clusters is known, even though the effectiveness is not guaranteed, the partitioning clustering is more computationally efficient than hierarchical clustering (Santosa et al. 2007).

Dimensionality reduction methods produce a low-dimensional projection of an originally high-dimensional data set. The most common methods used are: Singular

Value Decomposition (SVD), Principal Component Analysis (PCA) and Correspondence Analysis (CA). SVD can provide the subspace of any desirable dimension preserving the maximum possible similarity between the original data set and its projection onto the subspace. If the data matrix is squared, then SVD reduces eigenvalue decomposition. The data reduction technique used is PCA.

The derived uncorrelated components are linear combinations of the original variables whereas the uncorrelated variables can be removed or considered as residual, without losing much information about the original data. With several algorithms it is possible to calculate the principal components that can generate the same results from the same dataset. Instead, results may be different if there is more than one possible transformation for the same maximum variation (Bensmail et al. 2007). CA is an exploratory data analysis technique providing a view of the data set as whole. One of the main advantage of this technique is that it allows simultaneous observation of data samples (represented by column of the data matrix) and data points (represented by row of the data matrix) in one low dimensional space. The only restriction of this technique is that all data values must be non-negative. Busygin and Pardalos (2007) show how the CA is computationally efficient, especially for oncology problems.

The application fields of microarray technology are: cancer research, immunology, gene mutation, dystrophies research, etc. In the field of computational bioinformatics, many researchers are, in particular, interested in the biclustering problem and its biological interpretation.

3.1 Biclustering

Biclustering, also called co-clustering, shares the concepts of samples and features with clustering and classification, but it can be supervised or unsupervised. The main difference is that biclustering analyzes not only properties of samples but also their attributes (or features). In other words, it is a distinct class of clustering algorithms which is able to classify samples and features (rows and columns of the data matrix) simultaneously. Especially for microarray experiments, a dataset is given as a rectangular matrix $A \in R^{n\times m}$, where each column represents a data sample (e.g., patient, DNA sequence) and each row represents a feature (e.g., gene). Therefore, data are given by the matrix $A = (a_{ij})_{m\times n}$, where a_{ij} is the expression of the ith feature in the jth sample. Busygin et al. (2007) gave the following definition of biclustering.

Definition 1 A biclustering of a data set is a collection of pairs of samples and features' subsets $B = ((S_1, F_1), (S_2, F_2), \ldots, (S_r, F_r))$ such that the collection $(S_1, S_2, \ldots, S_r)$ forms a partition of the set of samples, and the collection $(F_1, F_2, \ldots, F_r)$ forms a partition of the set of features. A pair (S_k, F_k) will be called a biclustering.

In this definition, the authors consider just the case of one-to-one correspondence between classes of samples and features, but there are some other biclustering methodologies that do not consider this case. A common way to represent biclustering is by utilizing *heatmaps*. This is a rectangular pseudocolored image in which every pixel corresponds to a data value of the matrix A. Hence, by solving the biclustering problem, we wish to permute the lines and the rows of the data matrix so that a strong check board pattern appears in the heatmap (Fig. 1).

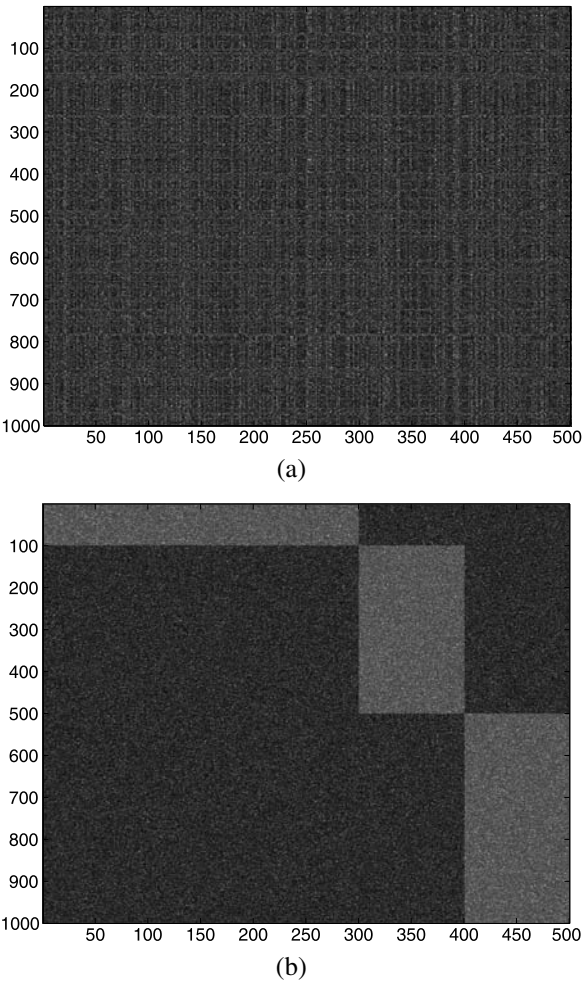

Fig. 1 Biclustering heatmaps: (**a**) a heatmap for a matrix before applying biclustering; (**b**) the same matrix permuted in such a way that the bicluster structure is clear

3.1.1 Biclustering analysis

Several papers have attempted to analyze biclustering based on their type, structure and algorithm (Tanay et al. 2005; Madeira and Oliveira 2004). In particular, Madeira and Oliveira (2004) have presented a review of these properties.

Bicluster type It is possible to identify four classes:

1. Biclusters with constant values (trivial case)
2. Biclusters with constant values on rows or columns (trivial case)
3. Biclusters with coherent values
4. Biclusters with coherent evolutions

The third type represents the case when each row or column can be obtained adding or multiplying each of the rows or columns by a constant value. Instead, algorithms

that identify biclusters with coherent evolutions, attempt to find a common behavior in the data matrix even if the numeric values within the bicluster do not match exactly.

Bicluster structure Based on the bicluster structure one can discriminate them in the following categories:

- *Exclusive row and column biclusters*. There is a block-diagonal structure after row and column reorder, where every row and every column in the matrix belongs exclusively to one of the K biclusters.
- *Nonoverlapping biclusters with checkerboard structure*. In this case, rows and columns may belong to more than one bicluster and assume a checkboard structure in the data matrix.
- *Exclusive rows biclusters*. The rows belong just to one bicluster, while column (conditions) can belong to several biclusters.
- *Exclusive columns biclusters*. The columns belong just to one bicluster and instead the rows (samples) can belong to more than one biclusters.
- *Nonoverlapping biclusters with tree structure*. A nonoverlapping bicluster does not allow a row or a column to belong in more than one biclusters.
- *Nonoverlapping nonexclusive biclusters*. It is the same as the previous case with the addition of the definition of exhaustive bicluster where every row and every column in the data matrix belongs to at least one bicluster. In the real case, it is more frequent to have some rows or columns that do not belong to any biclusters at all and that the bicluster overlap in some positions.
- *Overlapping biclusters with hierarchical structure*. It means that there are some biclusters separated or one that includes the others.
- *Arbitrarily positioned overlapping biclusters*. It is a general structure with K possibly overlapping biclusters without considering any order for its rows or columns.

3.1.1.1 Biclustering methods and algorithms There exist different algorithmic approaches for solving the biclustering problem and there are several survey papers on this topic (Fan et al. 2010; Xanthopoulos et al. 2010; Busygin and Pardalos 2007; Madeira and Oliveira 2004; Tanay et al. 2005). In the following section, we report some of the most prominent methods.

Node-deletion algorithm Cheng and Church were the first to introduce biclustering applied to gene expression data (Cheng and Church 2000; Tanay et al. 2005). The maximum biclustering problem is NP-hard but they suggest a greedy heuristic that quickly converges to a local maximal submatrix with a score smaller than a certain threshold δ. The algorithm is based on the minimization of the mean-squared residue score. The mean of the ith row of the sample cluster S_k is

$$\mu_{ik}^{r} = \frac{1}{|S_k|} \sum_{j \in S_k} a_{ij}, \tag{6}$$

whereas the mean of the jth column in the feature cluster F_k is

$$\mu_{jk}^{c} = \frac{1}{|F_k|} \sum_{i \in F_k} a_{ij}. \tag{7}$$

The mean value of the bicluster (S_k, F_k) is

$$\mu_k = \frac{\sum_{i \in F_k} \sum_{j \in S_k} a_{ij}}{|F_k||S_k|}. \tag{8}$$

Cheng and Church have proposed to measure the mean-squared residue score

$$\sum_{i \in F_k} \sum_{j \in S_k} \left(a_{ij} - \mu^r_{ik} - \mu^c_{jk} + \mu_k\right)^2. \tag{9}$$

Hence, they propose a two-step strategy to find the largest bicluster in A: removing rows and columns (superior to a given threshold) from the full matrix and adding the removed rows and columns considering a scoring scheme. More precisely, they have defined a score for each candidate bicluster and developed a heuristic to optimize this score function and its constraints. This algorithm discovers one bicluster at a time. An improved version of the Cheng and Church's algorithm is proposed by Yang et al. (2003). They propose the FLOC (FLexible Overlapped biClustering) algorithm that is able to determine simultaneously the biclusters and to achieve a solution quickly when some data are missed.

Spectral biclustering This algorithm, which is based on a bipartite graph model is proposed by Dhillon (2001). The graph is $G(V, E)$ where V is the set of vertices and $E \subseteq V \times V$. Kluger et al. (2003) have proposed a similar spectral biclustering applied for microarray data based on SVD. The edge (i.e., level of overexpression or underexpression of the gene under certain conditions) between the vertex corresponding to the feature $i = 1, \ldots, m$ and the vertex corresponding to the sample $j = 1, \ldots, n$ has a weight a_{ij}, whereas there are no connections (edges) between the vertices representing the feature (i.e., experimental condition) and between the one representing the samples (i.e., gene). The weighted adjacency matrix of the bipartite graph $G(V, E)$ is

$$W = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}. \tag{10}$$

Using graph partitioning, for instance, $V = V_1 \cup V_2$, the cut is the sum of the weights of the edges between these two partitions

$$\text{cut}(V_1, V_2) = \sum_{i \in V_1, j \in V_2} w_{ij}, \tag{11}$$

where $W = (w_{ij})_{n \times n}$ is the adjacency or affinity matrix constructed from data matrix A.

The most commonly used kinds of cuts are: the ratio (Dhillon 2001; Kluger et al. 2003), normalized (Kluger et al. 2003; Shi and Malik 2000), minimax (Zha et al. 2001), ICA (Rege et al. 2008), etc. The objective is to find a cluster of minimal weight, in other words, to minimize the cut between the disjoint set

$$\min\left(\frac{\text{cut}(V_1, V_2)}{\text{weight}(V_1)} + \frac{\text{cut}(V_1, V_2)}{\text{weight}(V_2)}\right), \tag{12}$$

where the weight function is $\text{weight}(V_1) = \sum_{i \in V_1} j \in V_1$.

Minimizing this expression is equivalent to solving a generalized eigenvalue problem subject to binary constraints which is *NP*-hard (Shi and Malik 2000). Relaxing the binary constraints leads to the solution given by the second smallest eigenvector of the Laplacian matrix $L := D - W$, where $D = \text{diag}(W \cdot 1^T)$, and 1 is the vector that has all the elements equal to 1.

Statistical algorithmic method for bicluster analysis The Statistical Algorithmic Method for Bicluster Analysis (SAMBA) is a biclustering algorithm that uses probabilistic modeling for performing simultaneous bicluster identification. Tanay et al. (2002) analyzed the performance of this algorithm on several gene expression datasets. SAMBA algorithm uses graph-theoretic techniques and probabilistic modeling. A bicluster corresponds to a subgraph $H = (U', V', E')$, where the weight is determined by the sum of the weights of edges $(u, v) \in E'$, and not edges $(u, v) \in \bar{E}' = (U' \times V')/E'$. The objective of SAMBA algorithm is to find the heaviest (largest weight) subgraph. To search heavy subgraphs, SAMBA uses a heuristic because the problem is NP-hard (Tanay et al. 2002) and it consists of two steps. In the first step, the bipartite graph is constructed and the weights of vertex pairs are computed. In the second, the algorithm looks for heavy subgraph around each vertex. Furthermore, SAMBA algorithm is able to analyze large datasets in minutes.

Consistent biclustering via fractional 0–1 programming A problem used especially for genomic and proteomic data analysis is the fractional 0–1 programming problem which is defined as

$$\max_{x \in \{0,1\}^m} f(x) = \sum_{j=1}^{n} \frac{\alpha_{j0} + \sum_{i=1}^{m} \alpha_{ji} x_i}{\beta_{j0} + \sum_{i=1}^{m} \beta_{ji} x_i}, \tag{13}$$

where $x_j \in \{0, 1\}$ and the denominator is positive. This problem is *NP*-hard (Kundakcioglu and Pardalos 2009). Busygin et al. (2005) defined a new class of fractional 0–1 programming problems where the fractional terms are not in the objective function but in the constraints. The main objective is to select the maximal possible number of features in order to lose minimal amount of information provided by the training set. One of the possible fractional 0–1 formulations based on biclustering criterion is

$$\begin{aligned} \max \quad & \sum_{i=1}^{m} x_i \\ \text{s.t.} \quad & \frac{\sum_{i=1}^{m} a_{ij} f_{i\hat{k}} x_i}{\sum_{i=1}^{m} f_{i\hat{k}} x_i} \geq (1 + t) \frac{\sum_{i=1}^{m} a_{ij} f_{ik} x_i}{\sum_{i=1}^{m} f_{ik} x_i}, \end{aligned}$$

where $x = (x_i)_{i=1,\ldots,m}$ is a vector of 0–1 variables. The variable x_i will be equal to 1 if the ith feature is selected and 0 otherwise. To solve this problem with fractional constraints, an approach which reduces the problem and which is similar to the one used to linearize problems with fractional 0–1 objective function is proposed (Wu 1997). Upon linearization the model is solved using standard linear mixed 0–1 programming solver (CPLEX). The parameter t in the above formulation is user defined (for details, see Wu 1997).

4 Data mining in computational neuroscience

In neuroscience, the electroencephalogram (EEG) is a popular low cost non-invasive examination for studying the human brain. It is extensively used in clinical practice as a diagnosis assisting tool and in research level for studying the response of the brain in various stimuli (e.g., event-related potentials ERP).

4.1 Epileptic seizure prediction

Optimization, among others, has found application in the emerging field of quantitative EEG analysis, and more precisely in understanding the dynamics of the brain. It has been suggested by many researchers (Milton and Jung 2003; Pardalos et al. 2004c) that human brain can be modeled as a dynamic system, meaning a system whose behavior is governed by a set of (unknown) differential equations. Based on this assumption, there have been many proposed algorithms for predicting state transition from normal into abnormal brain condition like epileptic seizures. For a comprehensive review of the existing literature about seizure prediction, we refer the reader to (Sackellares 2008; Mormann et al. 2007).

Usually, for quantification of brain state transitions, many researchers employ mathematical tools derived from chaos theory (Pardalos et al. 2003a, 2003b, 2004a). One of the most prominent tools is based on the so-called maximum Lyapunov exponent ($L_{\max}$). An approximation of $L_{\max}$, that is used especially when the signals are non-stationary, is the short term maximum Lyapunov exponent ($STL_{\max}$). Algorithms for efficient calculation of $STL_{\max}$ have been proposed in the literature (Sano and Sawada 1985; Wolf et al. 1985) and more recently in Pardalos and Yatsenko (2006). In common practice, the signal is divided into time windows, so that the properties of stationarity are valid, and for every time window $STL_{\max}$ is computed. Then the window slides in time and a new $STL_{\max}$ time series is generated.

Based on $STL_{\max}$, Iasemidis (1990) proposed T_{index} as a measure of nonlinear divergence between electrode sites. T_{index} between electrodes i and j is defined as

$$T_{\text{index}}(t) = \sqrt{N}\frac{|\overline{D}^t_{i,j}|}{\sigma^t_{i,j}}, \tag{14}$$

where $\overline{D}^t_{i,j}$ is the mean value of the differences of the $STL_{\max}$ values between electrode sites i and j over the time window t, $\sigma^t_{i,j}$ is the corresponding standard deviation, and N is the number of points contained in interval t. T_{index} is a statistic that follows asymptotically the Students t distribution with $N-1$ degrees of freedom. Therefore, it defines some measure of nonlinear coupling between all electrode pairs. It has been observed that this T_{index} drops significantly approximately 10 to 20 minutes before an epileptic seizure. T_{index} analysis produces one time series for every electrode pair (typically the total number of electrodes is around 30).

An optimization problem arises when one wants to select a set of k optimal electrodes based on some training dataset and use them for a seizure warning algorithm (versus selecting k electrodes randomly). In Pardalos et al. (2004a), the authors suggested a quadratic and multiquadratic approach for solving the *Electrode Selection*

Problem (*ESP*): If we denote with x the binary vector, where $x_i = 1$ if the ith electrode is selected and 0 otherwise, and A is the matrix with all the pairwise T_{index} values over a time window t, then the ESP can be formulated as

$$\begin{aligned} \min \quad & x^T A x \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = k, \\ & x \in \{0,1\}^n. \end{aligned}$$

Here, $A \in \mathbb{R}^{n \times n}$ is the matrix with the t-index values for all electrode pairs. The variable x is the binary vector of length n (x_i is 1 if electrode i is selected, 0 otherwise). k is the number of critical electrode sites that we want to select. This problem was solved using three alternative approaches: (i) a branch-and-bound scheme, (ii) a linearization technique for converting Quadratic Integer Programming (QIP) into a linear Integer Programming (IP) and (iii) by considering the corresponding system of optimality conditions (KKT).

Clinical and quantitative findings suggested that after an epileptic seizure dynamic resetting of the brain occurs, meaning that T_{index} reaches an increased value short after the seizure onset (Shiau et al. 2000). In order to embody this clinical finding into the electrode selection problem formulation, one additional quadratic constraint was introduced to the model:

$$\begin{aligned} \min \quad & x^T A x \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i = k, \\ & x^T C x \geq T_\alpha k(k-1), \\ & x \in \{0,1\}^n, \end{aligned} \tag{15}$$

where C is the matrix that contains all the pairwise T_{index} values for a time window of length ten minutes after the epileptic seizure onset and T_α is a critical T_{index} value that would reject the null hypothesis "brain sites i and j have the same $STL_{\max}$ values at time t". Although in Pardalos et al. (2004a) both problems were proved to be NP-hard practical instances of size ≈ 30, they were solved in a fair amount of time.

Two approaches were proposed for solving this problem: one using conventional linearization of the nonlinear constraints, and one linearization approach based on KKT optimality conditions. First, the KKT system was formulated and then the nonlinear constraints of the system were linearized. More precisely, it was proved that solving (15) is equivalent to solving the following integer linear program:

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} s_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} x_i - k = 0, \end{aligned}$$

$$-\sum_{j=1}^{n}\alpha_{ij}x_j + s_i - y_i = 0, \quad i = 1,\ldots,n, \tag{16}$$

$$y - M(1 - x_i) \le 0, \quad i = 1,\ldots,n, \tag{17}$$

$$h_i - Mx_i \le 0, \quad i = 1,\ldots,n, \tag{18}$$

$$-\sum_{j=1}^{n} d_{ij}x_j + h_i \le 0, \quad i = 1,\ldots,n, \tag{19}$$

$$-\sum_{j=1}^{n} h_i \ge T_\alpha k(k-1), \quad i = 1,\ldots,n, \tag{20}$$

$$x \in \{0,1\}^n, \quad s_i, y_i, h_i \ge 0, \quad i, j = 1,\ldots,n, \tag{21}$$

where $M = \|A\|_\infty$ is used to linearize the equations derived from the KKT system. The advantage of this formulation is that the number of additional constraints increase linearly with the size of the problem (with respect to the number of electrodes n) whereas a conventional linearization approach produces $O(n^2)$ additional constraints. As noticed in Chaovalitwongse (2003), efficiency for solving this problem will be crucial in future clinical setups when the electrode selection problem would need to be solved for larger n (future microelectrode technology would require $n \approx 300$).

4.2 Inverse brain source localization

One very important problem in quantitative neuroscience is to determine the current sources inside the brain given the scalp EEG recordings. This problem is generally a difficult problem that can accept multiple solutions. Several algorithms derive solutions for the problem by making specific considerations and assumptions. One challenging aspect of this problem is that the mathematical assumptions should comply with the biophysiological constraints. Generally, there are two models used for solving this problem: the underdetermined and the overdetermined. In the following sections, we will give a brief overview of the most prominent modeling for the inverse EEG problem.

4.2.1 Underdetermined model

In this modeling, which is also known as nonparametric or distributed source model, one is given the EEG recordings that can be denoted as a matrix $F \in \mathbb{R}^{n\times m}$ (n recording channels and m time samples). The problem is to determine the current distribution matrix $J \in \mathbb{R}^{3p\times m}$ over a dense grid of p space points through m time samples.

The human head conductance is modeled by a matrix $K \in \mathbb{R}^{n\times 3p}$ which is called *lead field matrix*. Determining a forward model is often referred as the *forward problem*. For an extensive review of solving the forward problem in EEG, we refer the reader to the following survey (Hallez et al. 2007).

Any solution of the inverse problem should satisfy the following equation:

$$F = K \cdot J. \tag{22}$$

This is a well known ill-posed problem meaning that we have more variables than equations, and because of that the problem can take many alternative solutions. One of the first optimization formulations was as a least-squares minimization problem:

$$\min_J \|F - KJ\|_F, \tag{23}$$

where $\|\cdot\|_F$ is the Frobenius norm. In this case, the estimate for the current source distribution will be

$$\hat{J}_{MNE} = \left(K^T K\right)^{-1} K^T F = K^{\dagger} F, \tag{24}$$

$K^{\dagger}$ is also referred as the Moore–Penrose pseudoinverse matrix. However, intuitively, the least-squares solution given by (24) is not very good in terms of neurophysiology because it tends to favor solution points close to the surface of the head. For this, several normalization approaches were proposed. For example, the weighted minimum norm formulation of the inverse problem is given by

$$\min_J \|F - WKJ\|_F, \tag{25}$$

where W is appropriate weighting matrix (usually constructed from the columns of the lead field matrix $W = \text{diag}(1^T K)$). In this case, the solution is given by

$$\hat{J}_{WMNE} = W^{-1} K^T \left[K W^{-1} K^T\right] F, \tag{26}$$

where $W = \Omega^2 \otimes I_3$ (by $\otimes$ we denote the Kronecker product), and Ω is constructed directly from the columns of the lead field matrix. Other pseudoinversing solutions consider a different weighting matrix (e.g., LORETA algorithm Pascual-Marqui 1999 embodies a Laplacian operator, or FOCUSS algorithm determines the weighing matrix iteratively Gorodnitsky et al. 1995). In general, such regularization methods converge to the solution by considering the first-order optimality conditions.

In general, there is a lot of interest from the biomedical community on the construction approach of a weighting matrix capable to embody realistic biophysical constraints. However, from the computational point of view, solutions can easily be estimated from explicit expressions. Nevertheless, the overdetermined formulation has been criticized for failing to embody biophysiological constraints, thus overdetermined modeling is preferred in most cases.

4.2.2 Overdetermined models

In the overdetermined models, we try to estimate a small number of current dipoles that optimally reconstruct the recorded voltage of the surface (EEG). In this category, some try to decompose the EEG recording signal into a small number of statistically independent (or uncorrelated) signals and then localize each one separately, or find the optimal configuration of the independent sources that reconstruct the surface

recording. The key assumption made is that the brain source signals are statistically independent with each other (something that might not be true in many cases).

One way to get an estimate for the number of uncorrelated signals of a mixture of time series is by looking at the eigenvalues of the variance–covariance matrix $E(FF^T)$ where $E(\cdot)$ denotes the expected value.

In other words, one can separate the data matrix into a sum of two terms as follows:

$$F = U\Lambda U^{-1} = U_s \Lambda_s U_s^{-1} + U_n \Lambda_n U_n^{-1} = F_s + F_n, \tag{27}$$

where the s subscript stands for *signal* and n for *noise*. This decomposition can be done according to some eigenvalue criterion. The data matrix that corresponds to the signal subspace is then processed using some higher order statistic techniques usually independent component analysis (ICA), and then the final independent components computed correspond to brain dipoles and are localized either separately or simultaneously. Separate localization of sources corresponds to greedy approach (first, localize the component that corresponds to the highest eigenvalue, etc.). This approach, although it is computationally efficient, suffers from all the drawbacks of greedy strategies. On the other side, simultaneous localization of multiple sources can be a difficult optimization problem for size larger than 3–4 dipoles. Multiple source localization of independent components with PCA preprocessing is discussed in Weinstein et al. (2000), Zhukov et al. (2000). For the multiple localization of sources, a multistart SIMPLEX optimization schema is considered.

Until today, the problem of simultaneously localizing statistical independent signals has been addressed with the use of optimization heuristics. Simulated annealing is considered in Miga et al. (2002) in several simulated examples under the presence of noise. On the other side, there is a significant amount of literature on Artificial Neural Networks (ANN) (Robert et al. 2002; Abeyratne et al. 1991; Sclabassi et al. 2001; Sun and Sclabassi 2000; Tun et al. 2000; Van Hoey et al. 2000). Their performance is strongly related to the training dataset even though ANN studies report low error percentages and robust behavior in noise environment. So, any application and evaluation of ANN should be extremely careful.

There is some relevant literature for application of genetic algorithms (McNay et al. 1996; Jiang et al. 2000) on the inverse problem. More precisely in Jiang et al. (2000), the authors consider a genetic algorithm, a simulated annealing approach and a tabu search approach in simulated model for localizing three dipoles. It is found that all three algorithms converge to global optimum if the computational resources are unlimited.

It is worth mentioning that the same methodology applied for EEG recordings is also applied when we have magnetoencephalographic recording (MEG). The advantage of MEG is that the magnetic signals are not distorted or absorbed from tissues or bones. On the other hand, the drawbacks of MEG are the high operational cost (shielded room is required) and constrained portability.

At this point, we would like to emphasize that, although the inverse source localization problem has attracted the interest of engineers and computational neuroscientists, most of the approaches lack sound computational complexity analysis. We identified a literature gap regarding the theoretical foundations of the problem and the study of the properties in an optimization framework.

5 Concluding remarks

In this paper, we presented some of the emerging applications of optimization and data mining related to radiation therapy planning for cancer treatment, genome data analysis and computational neuroscience. All the aforementioned areas have still many interesting applications for optimization tools, and there is a lot of room for experimentation and modeling. Although the amount of literature around these problems is constantly increasing, there is still scope for improvement. Most problems discussed in this paper are known to be NP-hard, meaning that even if we have some algorithms that give near optimal solution, one cannot guarantee that these algorithmic schemes will perform well in all problem instances.

Acknowledgements This work was funded by Air Force and NSF grants. Vera Tomaino was supported by the Department of Experimental Medicine and Clinic, University Magna Græcia of Catanzaro, Italy. The authors would like to acknowledge Ehsan Salari for his valuable comments and the feedback on the manuscript and also Sibel Bilge Sonuç for her proofreading and useful suggestions for the article.

References

Abeyratne R, Kinouchi Y, Oki H, Okada J, Shichijo F, Matsumoto K (1991) Artificial neural networks for source localization in the human brain. Brain Topogr 4(1):3–21

Aleman DM, Glaser D, Romeijn HE, Dempsey JF (2007) A primal-dual interior point algorithm for fluence map optimization in intensity modulated radiation therapy treatment planning. Technical report, Gainesville, Florida: Department of Industrial and Systems Engineering, University of Florida

Aleman DM, Romeijin HE, Dempsey JF (2008) Beam orientation optimization methods in intensity modulated radiation therapy treatment planning. In: Optimization in medicine and biology. Auerbach Publications, pp 223–251

Alves CJS, Pardalos PM, Vicente LN (eds) (2008) Optimization in medicine. Springer optimization and its applications, vol 12. Springer, Berlin

Bensmail H, Semmes OJ, Haoudi A (2007) Clustering proteomics data using Bayesian principal component analysis clustering proteomics data using Bayesian principal component analysis. In: Data mining in biomedicine. Springer optimization and its applications, vol 7. Springer, Berlin, pp 339–362

Busygin S, Pardalos PM (2007) Exploring microarray data with correspondence analysis. In: Data mining in biomedicine. Springer optimization and its applications, vol 7. Springer, Berlin, pp 25–37

Busygin S, Prokopyev OA, Pardalos PM (2005) Feature selection for consistent biclustering via fractional 0–1 programming. J Comb Optim 10:7–21

Busygin S, Prokopyev O, Pardalos PM (2007) Biclustering in data mining. Comput Oper Res 35(9):2964–2987

Chaovalitwongse W (2003) Optimization and dynamical approaches in nonlinear time series analysis with application in bioengineering. PhD thesis, University of Florida

Cheng Y, Church GM (2000) Biclustering of expression data. In: Proceedings of the eighth international conference on intelligent system for molecular biology, pp 93–103

Dhillon IS (2001) Co-clustering documents and words using bipartite spectral graph partitioning. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, pp 269–274

Fan N, Boyko N, Pardalos PM (2010) Recent advances of data biclustering with application in computational neuroscience. Springer, Berlin (in press)

Floudas C, Pardalos PM (eds) (2000) Optimization in computational chemistry and molecular biology. Kluwer Academic, Dordrecht

Gorodnitsky IF, George JS, Rao BD (1995) Neuromagnetic source imaging with FOCUSS: a recursive weighted minimum norm algorithm. Electroencephalogr Clin Neurophysiol, pp 231–251

Intensity Modulated Radiation Therapy Collaborative Working Group (2001) Intensity modulated radiotherapy: current status and issues of interest. Int J Radiat Oncol Biol Phys 51:880–914

Hallez H, Vanrumste B, Grech R, Muscat J, De Clercq W, Vergult A, D'Asseler Y, Camilleri KP, Fabri SG, Van Huffel S, Lemahieu I (2007) Review on solving the forward problem in EEG source analysis. J NeuroEng Rehabil 4(46)

Iasemidis LD (1990) On the dynamics of the human brain in temporal lobe epilepsy. PhD thesis, University of Michigan, Ann Arbor

Jiang T, Li X, Kruggel F (2000) Global optimization approaches to MEG source localization. pp 223–230

Kalinowski T (2008) Multileaf collimator shape matrix decomposition. In: Optimization in medicine and biology. Auerbach Publications, pp 253–286

Kamath S, Sahni S, Palta J, Ranka S, Lim J (2009) Algorithms for sequencing multileaf collimator. In: Handbook of optimization in medicine. Springer optimization and its applications, vol 26. Springer, Berlin, pp 223–242

Kluger Y, Basri R, Chang JT, Gerstein M (2003) Spectral biclustering of microarray data: coclustering genes and conditions. Genome Res 13:703–716

Küfer KH, Monz M, Scherrer A, Süss P, Alonso F, Azizi Sultan AS, Bortfeld T, Thieke C (2009) Multicriteria optimization in intensity modulated radiotherapy planning. In: Handbook of optimization in medicine. Springer optimization and its applications, vol 26. Springer, Berlin, pp 123–167

Kundakcioglu OE, Pardalos PM (2009) The complexity of feature selection for consistent biclustering. In: Chaovalitwongse, WA, Butenko, S, Pardalos, PM (eds) Clustering challenges in biological networks, pp 257–266

Legras J, Legras B, Lambert J (1982) Software for linear and non-linear optimization in external radiotherapy. Comput Programs Biomed 15:223–242

Lim GJ (2008) Introduction to radiation therapy planning optimization. In: Optimization in medicine and biology. Auerbach Publications, pp 197–221

Lim GJ (2009) Optimization models and computational approaches for three-dimensional conformal radiation treatment planning. In: Handbook of optimization in medicine. Springer optimization and its applications, vol 26. Springer, Berlin, pp 53–81

Lim GJ, Lee EK (eds) (2008) Optimization in medicine and biology. Auerbach Publications

Lim GJ, Ferris MC, Wright SJ, Shepard DM, Earl MA (2007) An optimization framework for conformal radiation treatment planning. INFORMS J Comput 19(3):366–380

Madeira SC, Oliveira AL (2004) Biclustering algorithms for biological data analysis: a survey. IEEE/ACM Trans Comput Biol Bioinform 1:24–45

McNay D, Michielssen E, Rogers RL, Taylor SA, Akhtari M, Sutherling WW (1996) Multiple source localization using genetic algorithms. J Neurosci Methods 64(2):163–172

Miga MI, Kerner TE, Darcey TM (2002) Source localization using a current-density minimization approach. Biomed Eng, IEEE Trans Biomed Eng 49(7):743–745

Milton J, Jung P (eds) (2003) Epilepsy as a dynamic disease. Springer, Berlin

Mondaini RP, Pardalos PM (eds) (2008) Mathematical modelling of biosystems, vol 102. Springer, Berlin

Mormann F, Andrzejak RG, Elger CE, Lehnertz K (2007) Seizure prediction: the long and winding road. Brain 130(2):314–333

Pardalos PM, Hansen P (eds) (2008) Data mining and mathematical programming. American Mathematical Society, Providence

Pardalos PM, Principe JC (eds) (2002) Biocomputing. Biocomputing, vol 1. Springer, Berlin

Pardalos PM, Romeijn HE (eds) (2009) Handbook of optimization in medicine. Springer optimization and its applications, vol 26. Springer, Berlin

Pardalos PM, Yatsenko VA (2006) Optimization approach to the estimation and control of Lyapunov exponents. J Optim Theory Appl 128:29–48

Pardalos PM, Iasemidis L, Shiau DS, Sackellares JC, Chaovalitwongse W (2003a) Prediction of human epileptic seizures based on optimization and phase changes of brain electrical activity. Optim Methods Softw 18(1):81–104

Pardalos PM, Iasemidis L, Shiau DS, Sackellares V, Yatsenko J, Chaovalitwongse W (2003b) Analysis of EGG data using optimization, statistics, and dynamical systems techniques. Comput Stat Data Anal 44:391–408

Pardalos PM, Chaovalitwongse W, Iasemidis LD, Sackellares JC, Shiau DS, Carney PR, Prokopyev OA, Yatsenko VA (2004a) Seizure warning algorithm based on optimization and nonlinear dynamics. Math Program 101(2):365–385

Pardalos PM, Sackellares C, Carney P, Iasemidis L (eds) (2004b) Quantitative neuroscience. Kluwer Academic, Dordrecht

Pardalos PM, Sackellares JC, Carney PR, Iasemidis LD (eds) (2004c) Quantitative neuroscience models, algorithms, diagnostics, and therapeutic applications. Kluwer Academic, Dordrecht

Pardalos PM, Boginski V, Vazakopoulos A (eds) (2007) Data mining in biomedicine. Springer optimization and its applications, vol 7. Springer, Berlin
Pascual-Marqui RD (1999) Review of methods for solving the EEG inverse problem. Int J Bioelectromagn 1:75–86
Pugachev A, Xing L (2001) Pseudo beam's-eye-view as applied to beam orientation selection in intensity-modulated radiation therapy. Int J Radiat Oncol, Biol, Phys 51:1361–1370
Reemtsen R, Albert M (2009) Continuous optimization of beamlet intensities for intensity modulated photon and proton radiotherapy. In: Handbook of optimization in medicine. Springer optimization and its applications, vol 26. Springer, Berlin, pp 83–122
Rege M, Dong M, Fotouhi F (2008) Bipartite isoperimetric graph partitioning for data co-clustering. Data Min Knowl Discov 16(3):276–312
Robert C, Gaudy JF, Limoge A (2002) Electroencephalogram processing using neural networks. Clin Neurophysiol 113(5):694–701
Romeijn HE, Dempsey JF (2008) Intensity modulated radiation therapy treatment plan optimization. TOP Off J Span Soc Stat Oper Res 16(2):215–243
Sackellares JC (2008) Seizure prediction. Epilepsy Curr 8(3):55–59
Sano M, Sawada Y (1985) Measurement of the Lyapunov spectrum from a chaotic time series. Phys Rev Lett 55:1082–1085
Santosa B, Conway T, Trafalis T (2007) A hybrid knowledge based-clustering multi-class SVM approach for genes expression analysis. In: Data mining in biomedicine. Springer optimization and its applications, vol 7. Springer, Berlin, pp 262–274
Sclabassi RJ, Sonmez M, Sun M (2001) EEG source localization: a neural network approach. Neurol Res 23(5):457–464
Seref O, Kundakcioglu OE, Pardalos PM (eds) (2008) Data mining, systems analysis, and optimization in biomedicine, vol 953. Springer, Berlin
Shi J, Malik J (2000) Normalized cuts and image segmentation. IEEE Trans Pattern Anal Mach Intell 22(8):888–905
Shiau DS, Luo Q, Gilmore SL, Roper SN, Pardalos PM, Sackellares JC, Iasemidis LD (2000) Epileptic seizures resetting revisited. Epilepsia 41(S7):208–209
Sun M, Sclabassi RJ (2000) The forward EEG solutions can be computed using artificial neural networks. IEEE Trans Biomed Eng 47(8):1044–1050
Tanay A, Sharan R, Shamir R (2002) Discovering statistically significant biclusters in gene expression data. Bioinformatics 18(Suppl 1):S136–S144
Tanay A, Sharan R, Shamir R (2005) Biclustering algorithms: a survey. In: Handbook of computational molecular biology. Chapman and Hall/CRC Press, London/Boca Raton, pp 26–1, 26–17
Tun AK, Lye NT, Guanglan Z, Abeyratne UR, Saratchandran P (2000) RBF networks for source localization in quantitative electrophysiology. Crit Rev Biomed Eng 28:463–472
Van Hoey G, De Clercq J, Vanrumste B, Walle R, Lemahieu I, D'Have' M, Boon P (2000) EEG dipole source localization using artificial neural networks. Phys Med Biol 45:997–1011
Weinstein DM, Zhukov L, Potts G (2000) Localization of multiple deep epileptic sources in a realistic head model via independent component analysis. School of computing technical report UUCS-2000-004, University of Utah
Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov exponents from a time series. Physica D 16:285–317
Wu TH (1997) A note on a global approach for general 0–1 fractional programming. Eur J Oper Res 101:220–223
Xanthopoulos P, Boyko N, Fan N, Pardalos PM (2010) Biclustering: algorithms and application in data mining. In: Wiley of operations research and management science (in press)
Yang J, Wang H, Wang W, Yu P (2003) Enhanced biclustering on expression data. In: Proceeding of the third IEEE conference on bioinformatics and bioengineering, pp 321–327
Zha H, He X, Ding C, Simon H, MGu (2001) Bipartite graph partitioning and data clustering. In Proceedings of the tenth international conference on Information and knowledge management, pp 25–32
Zhukov L, Weinstein D, Johnson CR (2000) Independent component analysis for EEG source localization in realistic head models. Proc IEEE Eng Med Biol Soc, 22nd Ann Int Conf 3(19):87–96