12-16-2011

# Nonparametric Copula Density Estimation in Sensor Networks

Leming Qu
*Boise State University*

Hao Chen
*Boise State University*

Yicheng Tu
*University of South Florida*

DOI: 10.1109/MSN.2011.50

# Nonparametric Copula Density Estimation in Sensor Networks

Leming Qu
Department of Mathematics
Boise State University
Boise, Idaho 83725-1555, USA
Email: lqu@boisestate.edu

Hao Chen
Department of Electrical & Computer Engineering
Boise State University
Boise, Idaho 83725, USA
Email: haochen@boisestate.edu

Yicheng Tu
Department of Computer Science & Engineering
University of South Florida
Tampa, Florida 33620, USA
Email: ytu@cse.usf.edu

***Abstract*—Statistical and machine learning is a fundamental task in sensor networks. Real world data almost always exhibit dependence among different features. Copulas are full measures of statistical dependence among random variables. Estimating the underlying copula density function from distributed data is an important aspect of statistical learning in sensor networks. With limited communication capacities or privacy concerns, centralization of the data is often impossible. By only collecting the ranks of the data observed by different sensors, we estimate and evaluate the copula density on an equally spaced grid after binning the standardized ranks at the fusion center. Without assuming any parametric forms of copula densities, we estimate them nonparametrically by maximum penalized likelihood estimation (MPLE) method with a Total Variation (TV) penalty. Linear equality and positivity constraints arise naturally as a consequence of marginal uniform densities of any copulas. Through local quadratic approximation to the likelihood function, the constrained TV-MPLE problem is cast as a sequence of corresponding quadratic optimization problems. A fast gradient based algorithm solves the constrained TV penalized quadratic optimization problem. Numerical experiments show that our algorithm can estimate the underlying copula density accurately.**

***Index Terms*—sensor network; dependence; copula; copula density estimation;**

## I. Introduction

Sensor networks have attracted considerable attention in the past two decades [1]. Distributed inference using sensor networks remains as an active research area due to its advantages such as increased reliability and greater coverage over centralized processing. Distributed inference refers to the decision making problem involving multiple distributed agents. The recent emergence of wireless sensor networks (WSN) has added many new dimensions to this classical inference problem. The limited capacity and other resource constraints make it imperative that each sensor node compress (often with very high ratio) their local observations before forwarding the data onto other sensors or fusion centers. Study of WSN has greatly enriched the theory of distributed inference, as evidenced by the rapidly increasing literature in recent years. Perhaps more importantly, the feasibility of having networked mobile and miniature sensor nodes has greatly broadened potential applications beyond sensing and surveillance. Examples include health care, environmental monitoring, and monitoring and diagnostics of complex systems. Many sensor network topologies are possible for the distributed inference problem, of which we focus on the parallel fusion topology which is the most popular and widely used one.

The observations by different sensors are often dependent. For simplicity, independence assumption is often made. The efforts to take the dependence into account is often carried out within a parametric modeling framework, in which the observed data are assumed to follow specific models such as a Gaussian probability distribution for random observations.

Copulas are full measures of statistical dependence among random variables. Understanding and quantifying dependence is an important, yet challenging, task in multivariate statistical modeling. In a linear, Gaussian world stochastic dependencies are captured by correlations. In more general settings, one often needs a complete specification of a joint distribution to have a complete knowledge of the dependence structure. The difficulty level of building such multivariate distributions can be greatly lowered if one uses a copula model to separate the marginal distributions from the dependence structure. Copula (otherwise known as dependence function) has emerged as a useful tool for modeling stochastic dependence. Sklar's theorem [2] is the theoretical foundation of the copula usage which states that a joint multivariate cumulative distribution function (CDF) equals the copula function of all univariate marginal CDFs. If all the univariate marginal CDFs are continuous, then the copula is unique. In other words, a copula is a multivariate CDF with standard uniform marginals. A copula density is the partial derivative of the copula, just as a joint multivariate probability density function (PDF) is the partial derivative of the joint CDF for continuous random variables. The name "copula" was chosen to emphasize the manner in which a copula "couples" a joint CDF to its univariate marginals. Some recent review papers on copulas include [3]–[7]. Some recent books on copulas include [8]–[10].

In the past two decades, copulas have been widely used in a variety of applied work, notably in finance and insurance. See [3], [8], [11]–[16] for example applications specific to finance and insurance. Some copula applications started to appear in signal and image processing recently. In [17], connections between Cohen-Posch theory of positive time-

DOI: 10.1109/MSN.2011.50

frequency distributions and copula theory were established. In [18], useful copula models for image classification were used in the frame of multidimensional mixture estimation arising in the segmentation of multicomponent images. In [19], the problem of detecting footsteps was considered where copulas were used to fuse acoustic and seismic measurements. In [20], Gaussian copula was used in the problem of tracking a colored object in video. Copula theory was used to detect changes between two remotely sensed images before and after the occurrence of an event in [21]. Generalized Gaussian copula was used for texture classification in [22]. A new divergence measure based on the copula density functions for image registration was explored in [23]. A possible link between copula and tomography was elaborated in [24]. In [25], a copula-based semi-parametric approach for footstep detection using seismic sensor networks was proposed. In [26], a novel approach for the fusion of correlated decisions to detect random signals under a distributed setting was proposed using the copula theory. In [27], a maximum-likelihood estimation based approach using copula functions was proposed to estimate the location of a source of random signals using a network of sensors. In [28], a parametric copula based framework for hypothesis testing using heterogeneous data was presented.

In these applications, parametric model assumptions were typically motivated by data or prior application-specific domain knowledge. However, when data is sparse or prior knowledge is vague, these parametric models may become questionable. Nonparametric methods are often desirable in such situations. Predd et al. [29] surveyed nonparametric distributed learning in WSN.

In what follows, we focus on the bivariate case only for simplicity. That is, we assume the network has two sensors, with each observing a specific random variable. For example, the first sensor records the temperatures in the surrounding, while the second sensor records the pressure in the same surrounding. The fusion center wishes to learn how the temperature and pressure are related to each other through the copula density. The proposed methodology is extendable to more than two dimensions.

The copula density estimation has been mostly studied in a parametric framework, whereby a bivariate copula density $c(u, v)$ is assumed to be a member of a copula family determined by a few parameters (for example, [30]). The parametric copula density estimation problem is then essentially reduced to estimate the few parameters that determine the copula. Choros et al. [31] provided a brief survey of parametric, semiparametric and nonparametric estimation procedures for copula models. We propose here to learn the bivariate copula density in the fusion center nonparametrically. For practitioners, nonparametric estimates could be used as the first step toward selecting the right parametric family.

Nonparametric estimation of copula and its density does not assume a specific parametric form for the copula and the marginals and thus provides great flexibility and generality. Nonparametric estimators of a bivariate copula density using kernels have been suggested by [32] and [33]. The advantage of kernel based copula density estimation is that it provides a smooth (differentiable) reconstruction of the copula function without putting any particular parametric a priori on the dependence structure between margins and without losing the usual parametric rate of convergence [33]. Kernel estimators have a severe drawback as they require a very large amount of data (page 195, [34]) and suffer from a corner bias. Nonparametric estimator of a copula using splines was proposed in [35] for a new class of copulas called linear B-spline copulas. Sancetta and Satchell [36] employed techniques based on Bernstein polynomials. Bernstein copula family belongs to the family of polynomial copulas [9] and can be used as an approximation to any copula. Nonparametric estimators of a copula density using wavelets were proposed in [37] [38] and [39]. Hall and Neumeyer [37] used a wavelet estimator to approximate a copula density. Genest et al. [38] used wavelet analysis to construct a rank-based estimator of a copula density. Autin et al. [39] dealt with the copula density estimation using wavelet methods by adaptive shrinkage procedures based on thresholding rules. These wavelet methods can better adapt to nonsmooth regions such as corners of a copula density.

What does a copula density $c(u, v)$ look like? In one extreme, for two independent random variables, $c(u, v)$ is a constant with value 1. When two random variables are dependent, $c(u, v)$ can be smooth, have sharp boundaries, or even be unbounded along boundaries. It is reasonable to assume that the total variation (TV) of $c(u, v)$, or at least its discrete version, is bounded. In practice, we often estimate and display the density in a finite grid. We propose a maximum penalized likelihood estimation (MPLE) with TV penalty method. This method is capable of capturing sharp changes in the target copula density, suffering less from edge effects when the copula density can be unbounded at boundaries in some statistically important cases, whereas conventional kernel or spline techniques have difficulties in nonsmooth regions. Our method preserves data privacy because each sensor is only required to send ranks of its individual records instead of the original observations to the fusion center.

The TV penalty based MPLE for copula density was proposed in [40], where the penalty term is the TV of the log density, and the unity requirement for a density function is imposed. However, the marginal unity, symmetry and positivity for a copula density are not enforced. In [41], the TV of the density is the penalty and the marginal unity and symmetry are enforced by linear equality constraints, but the positivity is not enforced. In fact, we are not aware of any method that explicitly imposes all the essential properties for a copula density. The main reason behind this is probably related to the difficulty of the induced high dimensional optimization problem. In this paper, we enforce the marginal unity and symmetry properties as linear equality constraints, and positivity property as linear inequality constraints for the discretized copula density. We solve the problem of minimizing penalized negative log likelihood with TV penalty subject to linear equality and inequality constraints through local quadratic approximation to the likelihood function first.

DOI: 10.1109/MSN.2011.50

The constrained TV-MPLE problem is then cast as a sequence of corresponding quadratic optimization problems. We apply a fast gradient based algorithm to solve the constrained TV penalized quadratic optimization problems. The effectiveness of our method is illustrated through numerical experiments.

The rest of the paper is organized as follows: In Section II, we formulate the problem. In Section III, we present the local quadratic approximation (LQA) algorithm, and in section IV show the experimental results. Finally, Section V concludes the paper.

## II. Problem Formulation

A bivariate copula density $c(u,v),\ \ [u,v] \in [0,1]^2$ can be regarded as the joint PDF of a bivariate standard uniform random variable $(U,V)$. Most copulas are exchangeable, thus implying $c(u,v)$ is symmetric. The $c(u,v)$ must satisfy the following four properties:

(P1) $c(u,v) \geq 0$, for $[u,v] \in [0,1]^2$ ;
(P2) $\int_0^1 c(u,v)du = 1$, for $0 \leq v \leq 1$;
(P3) $\int_0^1 c(u,v)dv = 1$, for $0 \leq u \leq 1$;
(P4) $c(u,v) = c(v,u)$.

Note that (P2) and (P4) implies (P3), so (P3) is redundant.

A bivariate copula $C(u,v)$ defined on the unit square $[0,1]^2$ is a bivariate CDF with univariate standard uniform margins:

$$C(u,v) = \int_0^u \int_0^v c(s,t)\mathrm{d}s\mathrm{d}t.$$

Sklar's Theorem ( [2]) states that the joint CDF $F(x,y)$ of a bivariate random variable $(X,Y)$ with marginal CDF $F_X(x)$ and $F_Y(y)$ can be written as $F(x,y) = C(F_X(x), F_Y(y))$, where copula $C$ is the joint CDF of $(U,V) = (F_X(X), F_Y(Y))$. This indicates a copula connects the marginal distributions to the joint distribution and justifies the use of copulas for building bivariate distributions.

Let $X_1,\ldots,X_n$ be a random sample from the unknown distribution $F_X$ that is observed at sensor 1. Let $Y_1,\ldots,Y_n$ be a random sample from the unknown distribution $F_Y$ that is observed at sensor 2. Further assume $(X_1,Y_1),\ldots,(X_n,Y_n)$ be a random sample from the unknown distribution $F$ of $(X,Y)$. We wish to estimate the copula density function $c(u,v)$.

When the two marginal distributions are continuous, the copula density $c(u,v)$ is the unique bivariate density of $(U,V) = (F_X(X), F_Y(Y))$ as implied by Sklar's theorem. As copulas are not directly observable, a nonparametric copula density estimator has to be formed in two stages: obtaining the observations for $(U,V)$ first and then estimating the copula density based on these observations.

In the first stage, the original data set $(X_i,Y_i)$ for $i = 1,\ldots,n$ is converted to $(\hat{U}_i,\hat{V}_i) = (\hat{F}_X(X_i), \hat{F}_Y(Y_i))$, where $\hat{F}_X$ and $\hat{F}_Y$ are conventional estimators of $F_X$ and $F_Y$. If models are available for the marginal distributions of $X$ and $Y$ but not for the joint distribution, one can use a technique such as maximum likelihood to estimate the marginal distribution functions. Otherwise, some nonparametric univariate CDF estimation methods or simply the empirical CDFs (ECDFs) can be used. When ECDFs are used as the marginal CDF estimators (e.g., in [38], [39]), $\hat{U}_i = \text{rank}(X_i)/n$ where rank($X_i$) is the rank of $X_i$ among $X_1,\ldots,X_n$ and $\hat{V}_i = \text{rank}(Y_i)/n$ where rank($Y_i$) is the rank of $Y_i$ among $Y_1,\ldots,Y_n$. Hence $\{(\hat{U}_i,\hat{V}_i)\}_{i=1}^n$ is nothing but the standardized ranks which are close substitutes for the unobservable pairs $(U_i,V_i) = (F_X(X_i), F_Y(Y_i))$ forming a random sample from the copula $C(u,v)$.

In the second stage, we estimate the copula density $c(u,v)$ based on the observations $\{(\hat{U}_i,\hat{V}_i)\}_{i=1}^n$.

Here we do not assume any parametric form for $c(u,v)$ and instead, obtain an estimate of it that satisfies properties (P1-P4) and is defined on a partition of the unit square. Specifically, the partition equally divides domain of $c(u,v)$, $[0,1]^2$, into $N = m^2$ rectangle cells with cell size $(1/m) \times (1/m)$ . A reasonable grid size is $64 \times 64$ (i.e., $m = 64$) for sample size $n = 2000$ and $m = 32$ for $n = 500$. A much finer discretization will increase problem size unnecessarily. In most numerical scheme, one fixes a grid resolution of $1/m$ much smaller than $2^{-J_n}$ with $J_n = \lfloor \frac{1}{2}\log_2(\frac{n}{\log n})\rfloor$ (page 207 of [39]).

Let us use $i,j = 1,\ldots,m$ to index all the $N$ cells of this grid. On each cell $(i,j)$, let $x_{ij}$ denote the constant estimate of $c(u,v)$ over the cell and set $p_{ij}$ to the number of observations $\{(\hat{U}_i,\hat{V}_i)\}_{i=1}^n$ falling in this cell.

A naive solution to $x_{ij}$ is $\hat{x}_{ij} = p_{ij}N/n$, which produces the 3D-histogram of the relative frequencies of the pseudo-observations $(\text{rank}(X_i)/n, \text{rank}(Y_i)/n)$ measured on the grids of the unit square. To illustrate this, we generated random samples of size $n = 2000$ from the Gaussian copula with parameter $\theta = 0.5$ and displayed the 3D-histograms in panel (c) of Fig. 1. The true Gaussian(0.5) copula density is plotted in panel (a) followed by the rank-rank plot in panel (b). The distinctive features of this copula are apparent as evidenced by the sharp corners, but the histogram is rather erratic and rough compared to the true copula density. Similar plots are shown in panels (a), (b), (c) in Figs. 2, 3 and 4 for Clayton(0.8), Frank(4) and Gumbel(1.25) copula density respectively.

The histograms obviously do not satisfy the marginal unity property (P2). The marginal integral of $c(u,v)$ can be approximated by the Riemann sum

$$\int_0^1 c(u,v)du \approx \frac{1}{m}\sum_{i=1}^m x_{ij} = 1, \quad j = 1,\ldots,m$$

and

$$\int_0^1 c(u,v)dv \approx \frac{1}{m}\sum_{j=1}^m x_{ij} = 1, \quad i = 1,\ldots,m.$$

The marginal unity (P2) implies

$$\begin{aligned} \sum_{i=1}^m x_{ij} &= m, \quad j = 1,\ldots,m, \text{ and} \\ \sum_{j=1}^m x_{ij} &= m, \quad i = 1,\ldots,m. \end{aligned}$$

DOI: 10.1109/MSN.2011.50

The purpose of this paper is to present a smoothed version of the 3D-histogram that practitioners could use as: (1)a graphical tool to spot the key features of a copula dependence structure such as skewness or heavy-tail behavior, (2)as a model selection tool to choose a particular parametric copula from families of copulas. In more technical terms, what will be proposed is a non-parametric (rank-based) estimator of the copula density

$$c(u,v) = \frac{\partial^2}{\partial u \partial v} C(u,v), \quad [u,v] \in [0,1]^2.$$

The approach described here is based on the maximum penalized likelihood estimation with a total variation penalty and by enforcing the copula density properties. That is, we estimate a copula density $c(u,v)$ as a $m \times m$ discrete image $x$ by solving:

$$\min_x \; T_\lambda(x) = -\sum_{i,j} p_{ij} \log x_{ij} + \lambda \; \text{TV}(x),$$

$$\text{such that} \quad \sum_{i=1}^m x_{ij} = m, \quad j = 1,\ldots,m, \text{ and}$$
$$x_{ij} = x_{ji}, \quad i,\; j = 1,\ldots,m, \text{ and}$$
$$x_{ij} > 0, \quad i,\; j = 1,\ldots,m.$$

where $\lambda$ is a smoothing parameter controlling the smoothness of the estimate.

As is customary, a discrete image $x = [x_{ij}]_{i,j=1}^m \in \mathbb{R}^{m\times m}$ will be dealt with as a vector in the usual Euclidean space $\mathbb{R}^N$ through the column stacking isometry $x_{ij} \leftrightarrow x_{i+mj}$. In the sequel, $x$ could mean either a 2D array or a 1D column vector depending on the context in which it appears. The linear equality constraints in the above minimization problem can be written in the form $Ax = b$ by forming the $m(m+1)/2 \times N$ sparse matrix $A$ and $m(m+1)/2$-vector $b$. The details of $A, b$ can be found in [41].

In a typical TV-based image restoration problem, TV($x$) is based on the first order finite difference which stems from the piecewise constant assumption of the underlying image $x$. A well-known drawback of the first order finite difference based TV regularized estimates is the staircase effect: the estimated values produced by TV regularization tend to cluster in patches [42], [43]. A copula density function $c(u,v)$ is continuous for $[u,v] \in [0,1]^2$, hence the first order finite difference of $x$ may not be sparse. But the higher order finite difference of $x$ are typically sparse. We choose the second order finite difference to define TV($x$). The even higher order finite difference based TV($x$) leads to unrealistic oscillatory solutions in our numerical experiments.

The $L_2$ norm and second order finite difference based discrete TV is

$$x \in \mathbb{R}^{m\times m}, \quad \text{TV}_2(x) =$$
$$\sum_{i,j=2}^{m-1} \sqrt{(x_{i+1,j} - 2x_{i,j} + x_{i-1,j})^2 + (x_{i,j+1} - 2x_{i,j} + x_{i,j-1})^2},$$

and the $L_1$-based TV is

$$x \in \mathbb{R}^{m\times m}, \quad \text{TV}_1(x) =$$
$$\sum_{i,j=2}^{m-1} |x_{i+1,j} - 2x_{i,j} + x_{i-1,j}| + |x_{i,j+1} - 2x_{i,j} + x_{i,j-1}|$$

where we set the (standard) reflexive boundary conditions

$$x_{m+2,j} \equiv x_{m+1,j} \equiv x_{m,j}, \quad j = 1,\ldots m;$$
$$x_{i,m+2} \equiv x_{i,m+1} \equiv x_{i,m}, \quad i = 1,\ldots,m.$$

The algorithms developed in this paper can be applied to both the $L_2$- and $L_1$-TV. Since the derivations and results for the $L_2$ and $L_1$ cases are very similar, to avoid repetition, all of our derivations will consider the $L_1$-TV.

Let $l(x) \equiv -\sum_{i,j=1}^m p_{ij} \log x_{ij}$, then $\nabla l(x) = -p./x$, where $\nabla$ denotes the gradient operator with respect to $x$ and ./ denotes element-wise division. The Hessian $H(x) = \partial^2 l(x)/\partial x^2$ is a diagonal matrix with diagonal vector $p./x^2$ The gradient and Hessian will be used in the optimization algorithm discussed in the next section.

Our proposed copula density estimate solves:

$$\min_x \{l(x) + \lambda \text{TV}(x)\}, \text{ such that } Ax = b, \quad x > 0. \qquad (1)$$

When $\lambda \to \infty$, the solution of (1) is $\hat{x} = 1$ leading to $\hat{c}(u,v) = 1$ which implies the indepence of $X$ and $Y$ for the density estimation and apparently over-smoothes the density for dependent case. When $\lambda = 0$, the solution of (1) is the histogram estimate which under-smoothes the density. For an appropriately chosen $\lambda$, the solution is a properly regularized estimate with the right amount of smoothness.

## III. Local Quadratic Approximation (LQA) Algorithm

To solve problem (1), we first approximate $l(x)$ locally by its quadratic expansion around $x^k$ at $k$th iteration:

$$l(x) \approx l(x^k) + (x - x^k)^T \nabla l(x^k) + \frac{1}{2}(x - x^k)^T H(x^k)(x - x^k).$$

For $k = 0, 1, \ldots$, we then solve the following problem

$$\min_x \left\{ x^T \nabla l(x^k) + \frac{1}{2}(x - x^k)^T H(x^k)(x - x^k) + \lambda \text{TV}(x) \right\},$$
$$\text{such that } Ax = b, \quad x > 0, \qquad (2)$$

until certain convergence criteria is met.

Problem (2) is a special case of the optimization problem with a composite objective function [44]:

$$\min_x \{F(x) \equiv f(x) + g(x)\}, \qquad (3)$$

with the following assumptions:

- $g(x) : \mathbb{R}^{m\times m} \to (-\infty, +\infty]$ is a proper closed convex function which is possibly nonsmooth;
- $f(x) : \mathbb{R}^{m\times m} \to (-\infty, \infty)$ is continuously differentiable with Lipschitz continuous gradient

$$||\nabla f(x) - \nabla f(y)|| \le L(f)||x - y|| \quad \forall x, y \in \mathbb{R}^{m\times m}$$

DOI: 10.1109/MSN.2011.50

where $||\cdot||$ denotes the standard Euclidean norm and $L(f)$ is the Lipschitz constant of $\nabla f$;

- problem (3) is solvable, i.e., $B_* := \arg\min_x F(x) \neq \emptyset$.

Problem (3) reduces to problem (2) by setting

$$f(x) \equiv x^T \nabla l(x^k) + \frac{1}{2}(x - x^k)^T H(x^k)(x - x^k), \quad (4)$$

$$g(x) \equiv \lambda \mathrm{TV}(x) + \delta_C(x) \quad (5)$$

where $C = \{x \in \mathbb{R}^N : Ax = b, \quad x > 0\}$ a closed convex set and $\delta_C$ being the indicator function on $C$. The Lipschitz constant of this specific $f(x)$ is the maximum of the Hessian diagonal vector, which is $L_k = \max\left\{p./(x^k)^2\right\}$.

To solve problem (3), Beck and Teboulle [44] [45] proposed a gradient-based algorithm which shares a remarkable simplicity together with a proven global rate of convergence which is significantly better than currently known gradient projections-based methods. The algorithm is termed MFISTA which stands for monotone fast iterative shrinkage/thresholding algorithm. The key idea is to adopt a quadratic separable approximation to $f(x)$ around the current estimate $y$:

$$f_Q(x) \equiv f(y) + (x - y)^T \nabla f(y) + \frac{1}{2\alpha}||x - y||^2,$$

for a given $\alpha > 0$. This approximation interpolates the first-derivative information of $f(x)$ and uses a simple diagonal Hessian approximation to the second-order term. Hence a quadratic separable approximation to $F(x)$ around the current estimate $y$ is

$$\begin{aligned} F_Q(x) &\equiv f_Q(x) + g(x) \\ &= f(y) + (x - y)^T \nabla f(y) + \frac{1}{2\alpha}||x - y||^2 + g(x). \end{aligned}$$

This quadratic approximation to $F(x)$ can also be interpreted as a regularization method with a quadratic proximal term that would measure the local error in the linear approximation, and also results in a well defined, i.e., a strongly convex approximate minimization problem for (3) [46]:

$$\begin{aligned} x^* &\equiv \arg\min_x F_Q(x) \\ &= \arg\min_x \left\{0.5||x - (y - \alpha\nabla f(y))||^2 + \alpha g(x)\right\}. \quad (6) \end{aligned}$$

Apparently, the $\alpha$ serves the role of step size. There are many data adaptive ways to search for step size, including backtracking line search algorithm (section 3.1 of [47]) and spectral gradient method [48]. Under the assumption of $f(x)$ having Lipschitz continuous gradient, it turns out that $\alpha = 1/L(f)$ is a good fixed step size.

Instead of solving problem (6) at the current iterate $y$, FISTA algorithm smartly solve the problem at the $y$ which is formed by a very specific linear combination of the previous two iterates.

For a given $z \in \mathbb{R}^{m\times m}$ and a scalar $\alpha > 0$, the proximal map of Moreau [49] [50] associated to a convex function $g(x)$ is defined by

$$\mathrm{prox}_g(z, \alpha) := \arg \min_{x\in\mathbb{R}^{m\times m}} \left\{0.5||x - z||^2 + \alpha g(x)\right\}. \quad (7)$$

The problem (6) is obviously a proximal map by setting $z \equiv y - \alpha\nabla f(y)$.

When the subproblem (7) is not solved accurately, the FISTA may diverge. To get rid off this trouble, the objective function $F(x)$ in (3) is forced to be non-increasing at each step, which leads to the monotone version of FISTA: MFISTA.

With the specific $g(x)$ in (5), the problem (7) becomes the constrained TV-based denoising problem:

$$\min_{x\in C} \left\{0.5||x - z||^2 + \alpha\lambda \mathrm{TV}(x)\right\}. \quad (8)$$

One of the intrinsic difficulties to solve this problem is the nonsmoothness of the TV function. This was overcome by a dual approach in [44] which followed [51]. The objective function for the dual problem is continuously differentiable and its Lipschitz constant has an analytical upper bound. Hence, the dual problem is solvable by a fast gradient projection algorithm based on FISTA. The dual problem requires a projection onto the linear equality and positivity constrained convex set which we discuss below.

### *A. Projection onto Linear Equality and Positivity Constrained Convex Set*

For a given $y \in \mathbb{R}^N$, this projection is to solve

$$\min_x ||x - y||^2, \text{ such that } Ax = b, x > 0. \quad (9)$$

By penalizing the square of the $L_2$ norm of the difference between $Ax$ and $b$, we obtain the following approximation to problem (9)

$$\min_x \left\{0.5\beta||Ax - b||^2 + 0.5||x - y||^2\right\}, \quad \text{s.t.} \quad x > 0 \quad (10)$$

with a sufficiently large penalty parameter $\beta$. This type of quadratic penalty approach can be traced back as early as [52] in 1943. It was discussed in section 17.1 in [47]. It is well known that the solution of (10) converges to that of (9) as $\beta \to \infty$. This quadratic penalty approach was popular in large-scale underdetermined linear equality constrained sparse recovery problems in recent years [48]. It was used in a new alternating minimization algorithm for total variation image reconstruction [53] as well.

We prefer to solve the following equivalent problem to (10):

$$\min_x \left\{0.5||Ax - b||^2 + 0.5\mu||x - y||^2\right\}, \text{ s.t } x > 0 \quad (11)$$

with a small penalty parameter $\mu$, because problem (11) is more stable than (10) in case when $A$ is ill-conditioned.

It is straightforward to apply FISTA algorithm to solve problem (11) by setting

$$\begin{aligned} f(x) = 0.5||Ax - b||^2,\ \nabla f(x) = A^T(A\beta - b), \\ L(f) = \mathrm{maxeig}(A^T A),\ g(x) = 0.5\mu||x - y||^2 + \delta_{\{x>0\}}(x). \end{aligned}$$

We note that the maximum eigen value of $A^T A$ as $L(f)$ should be precomputed outside the main loop and be passed to FISTA

DOI: 10.1109/MSN.2011.50

to avoid redudant computation . The closed-form solution to the sub-problem (7) in this case is:

$$\begin{aligned}\text{prox}_g(z,\alpha) :&= \arg\min_{x>0} \{||x-z||^2 + \alpha\mu||x-y||^2\} \\ &= P_{\{x>0\}}\left(\frac{z+\alpha\mu y}{1+\alpha\mu}\right),\end{aligned}$$

where $P_{\{x>0\}}(\cdot)$ is simply the positivity projection operator.

## IV. Simulations

We conduct simulation studies designed to demonstrate the effectiveness of the TV-MPLE subject to linear equality and positivity constraints for copula density estimation.

The stopping criterion for LQA in the main algorithm was $||x^{k+1} - x^k||/||x^k|| <= 10^{-6}$ or total number of iterations reaching 10.

In the simulation, the marginal CDFs $F_X$ and $F_Y$ were estimated by ECDFs. This amounts to use the standardized ranks of the sample $\{(X_i, Y_i)\}_{i=1}^n$ as estimates of $\{(U_i, V_i)\}_{i=1}^n$ (remind that $U_i = F_X(X_i)$ and $V_i = F_Y(Y_i)$). The CDF of a continuous random variable is continuous and strictly increasing within its domain, which implies that the ranks of $X_i$'s are the same as the ranks of $U_i$'s, so are the ranks of $Y_i$'s and those of $V_i$'s. Therefore it is unnecessary to explicitly specify the $F_X$ and $F_Y$ in our simulation for copula density estimation. One can first generate $\{(U_i, V_i)\}_{i=1}^n$ from an underlying copula density $c(u, v)$, then use their standardized ranks as their estimates.

We tested four parametric families of copulas: Gaussian, Clayton, Frank and the Gumbel families. For each copula model, independent and identically distributed (i.i.d.) standard uniform bivariate random variables $\{(U_i, V_i)\}_{i=1}^n$ were generated from the specified copula with parameter $\theta$ using MATLAB's *copularnd*() function. That was, $\{U_i\}_{i=1}^n$ was a sample from a Uniform(0,1) distribution, and so was the $\{V_i\}_{i=1}^n$. The joint pdf of $(U, V)$ was the specified copula density $c(u, v)$ with parameter $\theta$. The sample sizes considered was $n = 2000$. The grid sizes used was $m = 64$.

Various error measures were evaluated over the equally spaced grid points within $[0, 1]^2$ where the copula densities were estimated. For one data set, the quality of an estimate $\hat{c}_\lambda(u, v)$ of the true copula density $c(u, v)$ was measured by an error measure $\text{Loss}(\hat{c}_\lambda, c)$, which is relative errors

$$RE_q(\lambda) = \frac{||\hat{c}_\lambda - c||_{N,q}}{||c||_{N,q}}, \quad \text{for } q = 1,\ 2. \tag{12}$$

The regularization parameter $\lambda$ was chosen from 10 equally spaced numbers in $[10^{-4}, 10^{-2}]$ in a log10 scale: $\lambda_1 > \lambda_2 > \ldots > \lambda_{10}$ with $\lambda_1 = 10^{-2}$ and $\lambda_{10} = 10^{-4}$. We started solving problem (1) with $\lambda = \lambda_1$ using the initial value $x_{init} = 1$. We then proceeded to solve problem (1) with $\lambda = \lambda_2$ and set the initial value for $x$ as the previous solution. This choice of initial value is the so-called warm start [53]. Finally, we select the best $\lambda$ out of $\{\lambda_1, \ldots, \lambda_{10}\}$. For the error measure $\text{Loss}(\hat{c}_\lambda, c)$, the best regularization parameter $\hat{\lambda}$ is the one which minimizes $\text{Loss}(\hat{c}_\lambda, c)$ and the best estimate is $\hat{c}_{\hat{\lambda}}$. All the best regularization parameters were found near the central portion of this grid.

Fig. 1 displays the surface plots of the estimated copula densities in panels (e) (f) for Gaussian(0.5) copula. For comparison, we computed a 2D kernel density estimate using the kde2D program [54] [55] as shown in panel (d). Obviously, there is an oversmoothing by KDE. The TV estimates catch the two peaks in the front and back corners well.

Figs. 2, 3 and 4 display the results for Clayton(0.8), Frank(4) and Gumbel(1.25) respectively. Again, The TV estimates are close to the truth.

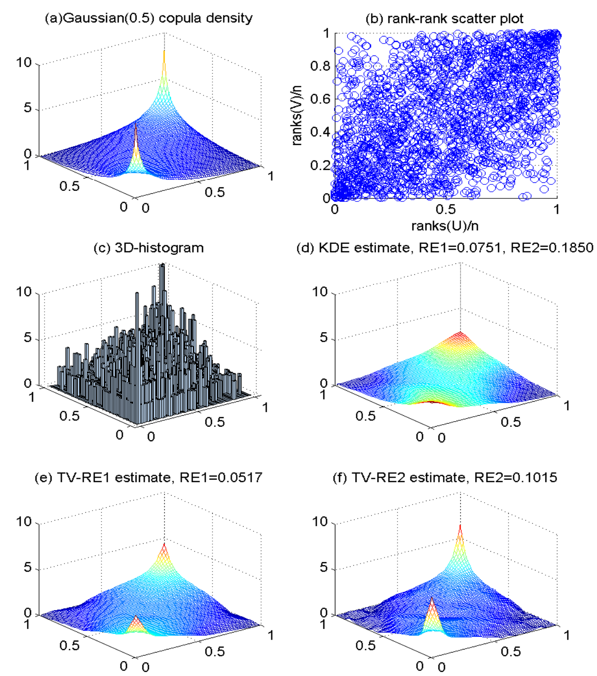


Fig. 1. True and estimated copula densities in a typical run of the case: Gaussian copula with $\theta = 0.5$, sample size $n = 2000$, grid size $m$ = 64.

Our TV-MPLE copula density estimate can serve the purpose to select a parametric copula from several parametric families. A parametric copula $c_\theta$ is wholly determined by its parameter $\theta$. The parameter $\theta$ can be estimated by classical parameter estimation methods such as maximum likelihood. We measure the distance between our nonparametric estimate $\hat{c}_\lambda$ and the parametric estimate $c_{\hat{\theta}}$ by their relative errors

$$RE_q(\hat{\theta}) = \frac{||\hat{c}_\lambda - c_{\hat{\theta}}||_{N,q}}{||c_{\hat{\theta}}||_{N,q}}, \quad \text{for } q = 1,\ 2, \infty.$$

The selected parametric copula is the one with the smallest $RE_q(\hat{\theta})$ among all parametric candidates.

A simulation study was to illustrate this model selection strategy. An i.i.d. standard uniform bivariate random sample $\{(U_i, V_i)\}_{i=1}^n$ with $n = 2000$ was generated from the Gaussian copula density with $\theta = 0.5$ . TV-MPLE estimate $\hat{c}_\lambda$ was constructed based on the data $\{(\hat{U}_i, \hat{V}_i)\}_{i=1}^n$ with grid size $m = 64$ and $\lambda$ selected by the best $\lambda$ in terms of $RE_2(\lambda)$.

DOI: 10.1109/MSN.2011.50

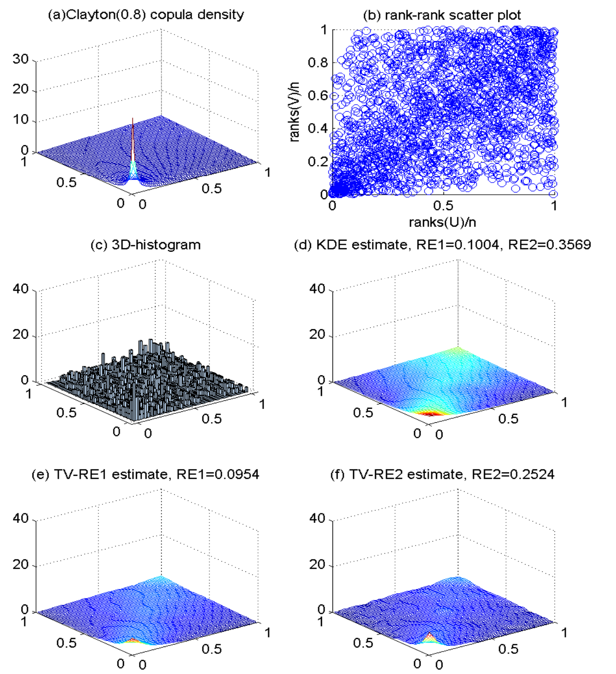


Fig. 2. True and estimated copula densities in a typical run of the case: Clayton copula with $\theta = 0.8$, sample size $n = 2000$, grid size $m$ = 64.

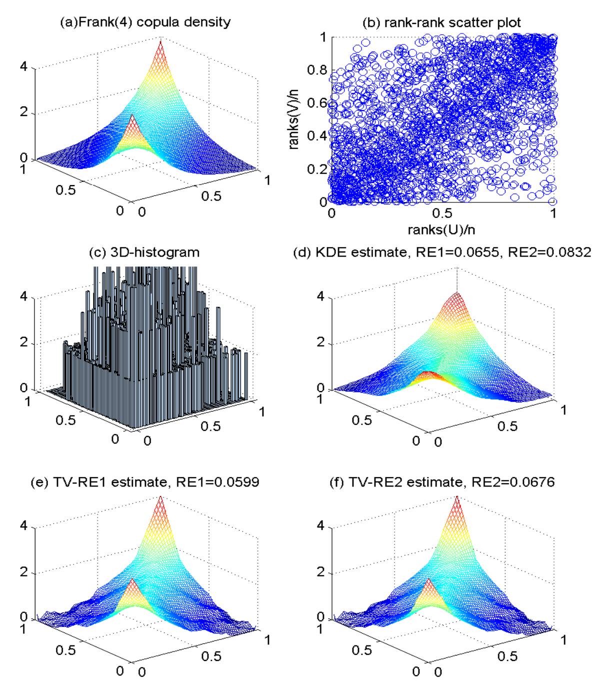


Fig. 3. True and estimated copula densities in a typical run of the case: Frank copula with $\theta = 4$, sample size $n = 2000$, grid size $m$ = 64.

The $\theta$ was estimated by the Canonical Maximum Likelihood (CML) method using MATLAB's *copulafit*() function. Table I reports the $RE_q(\hat{\theta})$ for 4 different candidates. TV-MPLE estimate is closest to the Gaussian estimate in terms of any

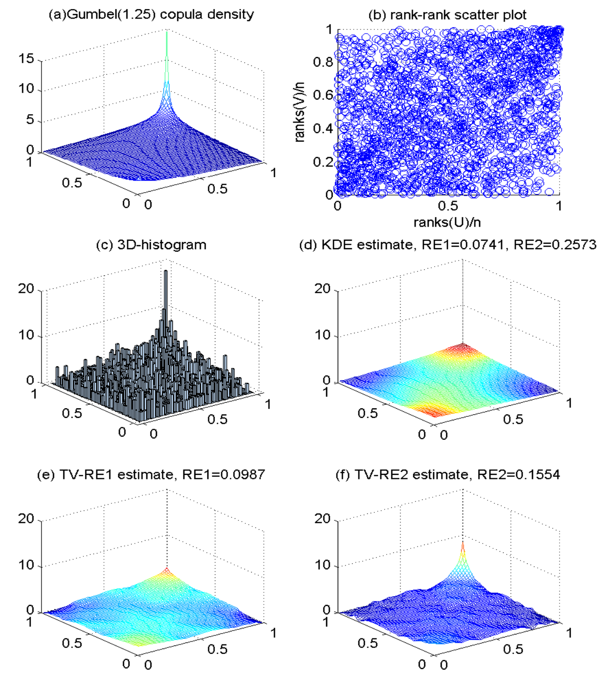


Fig. 4. True and estimated copula densities in a typical run of the case: Gumbel copula with $\theta = 1.25$, sample size $n = 2000$, grid size $m$ = 64.

relative errors. We correctly select the Gaussian model among four parametric families considered.

TABLE I
RELATIVE ERROR $RE_q(\hat{\theta})$ FOR THE SIMULATED DATA FROM GAUSSIAN(0.5) WITH $n = 2000$, $m = 64$

| Parametric Estimate | $RE_1(\hat{\theta})$ | $RE_2(\hat{\theta})$ | $RE_\infty(\hat{\theta})$ |
|---|---|---|---|
| Gaussian | 0.0527 | 0.0917 | 0.3997 |
| Clayton | 0.1621 | 0.3406 | 0.7605 |
| Frank | 0.0933 | 0.1270 | 0.5889 |
| Gumbel | 0.7567 | 0.7869 | 0.9468 |

## V. CONCLUDING REMARKS

In a distributed sensor network, working with rank data only in the processing center, we presented a TV penalized maximum likelihood copula density estimate subject to the constraints that the marginal distributions are standard uniforms. The linear equality and positivity constrained TV regularized MPLE problem is solved by a local quadratic approximation algorithm in the main iteration. The sub-problem of constrained quadratic programming with TV penalty is solved by MFISTA. The resulting nonparametric copula density estimate captures the salient features of the underlying copula density. The data adaptive choice of the regularization parameter $\lambda$ will be implemented in the future.

## REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, 2002.

DOI: 10.1109/MSN.2011.50


[2] A. Sklar, *Fonctions de répartition à n dimensions et leurs marges*, Publ. Inst. Stat. Univ. Paris, 1959.

[3] P. Embrechts, F. Lindskog, and A. McNeil, "Modelling dependence with copulas and applications to risk management," in *Handbook of Heavy Tailed Distributions in Finance*, S. T. Rachev, Ed. North-Holland, Amserdam: Elsevier, 2003, ch. 8, pp. 329–384.

[4] N. Kolve, U. dos Anjos, and B. Mendes, "Copulas: a review and recent developments," *Stochastic Models*, vol. 22, pp. 617–660, 2006.

[5] T. Mikosch, "Copulas: tales and facts," *Extremes*, vol. 9, pp. 3–20, 2006.

[6] P. Embrechts, "Copulas: a personal view," *Journal of Risk and Insurance*, vol. 76, pp. 639–650, 2009.

[7] A. P., "Copula-based models for financial time series," in *Handbook of Financial Time Series*, T. Andersen, R. Davis, J.-P. Kreiss, and T. Mikosch, Eds. Verlag: Springer, 2009, pp. 767–785.

[8] U. Cherubini, E. Luciano, and W. Vecchiato, *Copula Methods in Finance*. New York: John Wiley & Sons, 2004.

[9] R. B. Nelsen, *An Introduction to Copulas*, 2nd ed., ser. Lecture Notes in Statistics. New York: Springer, 2006.

[10] P. K. Trivedi and D. M. Zimmer, *Copula Modeling: An Introduction for Practitioners*. Hanover, Mass.: Now Publishers, 2007.

[11] E. W. Frees and E. A. Valdez, "Understanding relationships using copulas," *North American Actuarial Journal*, vol. 2, pp. 1–25, 1998.

[12] P. Embrechts, A. McNeil, and D. Straumann, "Correlation and dependence in risk management: Properties and pitfalls," in *RISK Management: Value at Risk and Beyond*, M. A. H. Dempster, Ed. Cambridge, U.K.: Cambridge University Press, 2002, pp. 176–223.

[13] Y. Malevergne and D. Sornette, "Testing the gaussian copula hypothesis for financial assets dependence," *Quant Fin*, vol. 3, p. 231250, 2003.

[14] M. Junker and A. May, "Measurement of aggregate risk with copulas," *Econ J*, vol. 8, pp. 428–454, 2005.

[15] J. Rodriguez, "Measuring financial contagion: a copula approach," *J Empir Fin*, vol. 41, pp. 401–423, 2007.

[16] S. Chen and S. Poon, "Modelling international stock market contagion using copula and risk appetite," 2007, http://ssrn.com/abstract=1024288.

[17] M. Davy and A. Doucet, "Copulas: A new insight into positive time-frequency distributions," *IEEE Signal Process. Lett.*, vol. 10, no. 7, pp. 215–218, Jul. 2003.

[18] N. Brunel, W. Pieczynski, and S. Derrode, "Copulas in vectorial hidden markov chains for multicomponent image segmentation," in *Proc. ICASSP*, 2005, pp. 717–720.

[19] S. G. Iyengar, P. K. Varshney, and T. Damarla, "On the detection of footsteps using acoustic and seismic sensing," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, 2007, p. 22482252.

[20] W. Ketchantang, S. Derrode, and L. Martin, "Pearson-based mixture model for color object tracking," *Mach. Vis. Appl.*, vol. 19, pp. 457–466, 2008.

[21] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.

[22] Y. Stitou, N. Lasmar, and Y. Berthoumieu, "Copula based multivariate gamma modeling for texture classification," in *Proc. ICASSP*, 2009, pp. 1045–1048.

[23] T. Durrani and X. Zeng, "Copula-based divergence measures and their use in image registration," in *EUSIPCO*, 2009.

[24] D.-B. Pougaza, A. Mohammad-Djafari, and J.-F. Bercher, "Link between copula and tomography," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2258–2264, Oct. 2010.

[25] A. Sundaresan, A. Subramanian, P. K. Varshney, and T. Damarla, "A copula-based semi-parametric approach for footstep detection using seismic sensor networks," in *Multisensor, Multisource Information Fusion: Architectures, Algorithms, and Applications*, J. Braun, Ed. SPIE, 2010, p. 77100C, vol. 7710.

[26] A. Sundaresan, P. Varshney, and N. Rao, "Copula-based fusion of correlated decisions," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 1, pp. 454–471, Jan. 2011.

[27] A. Sundaresan and P. Varshney, "Location estimation of a random signal source based on correlated sensor observations," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 787 –799, feb. 2011.

[28] S. Iyengar, P. Varshney, and T. Damarla, "A parametric copula-based framework for hypothesis testing using heterogeneous data," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2308–2319, May 2011.

[29] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Process. Mag.*, pp. 56–69, Jul. 2006.

[30] J. H. Shih and T. A. Louis, "Inferences on the association parameter in copula models for bivariate survival data," *Biometrics*, vol. 51, pp. 1384–1399, 1995.

[31] B. Choros, R. Ibragimov, and E. Permiakova, "Copula estimation," in *Copula Theory and Its Applications, Lecture Notes in Statistics, V198*, Springer, 2010, pp. 77–91.

[32] I. Gijbels and J. Mielniczuk, "Estimating the density of a copula function," *Communications in Statistics - Theory and Methods*, vol. 19, pp. 445–464, 1990.

[33] J. Fermanian and O. Scaillet, "Nonparametric estimation of copulas for time series," vol. 5, no. 4, pp. 25–54, 2003.

[34] Y. Malevergne and D. Sornette, *Extreme Financial Risks*. Heidelberg: Springer, 2006.

[35] X. Shen, Y. Zhu, and L. Song, "Linear B-spline copulas with applications to nonparametric estimation of copulas," *Computational Statistics & Data Analysis*, vol. 52, pp. 3806–3819, 2008.

[36] A. Sancetta and S. Satchell, "The Bernstein copula and its applications to modeling and approximations of multivariate distributions," *Econometric Theory*, vol. 20, pp. 535–562, 2004.

[37] P. Hall and N. Neumeyer, "Estimating a bivariate density when there are extra data on one or both components," *Biometrika*, vol. 93, pp. 439–450, 2006.

[38] C. Genest, E. Masiello, and K. Tribouley, "Estimating copula densities through wavelets," vol. 44, no. 2, pp. 170–181, Apr. 2009.

[39] F. Autin, E. L. Pennecb, and K. Tribouley, "Thresholding methods to estimate the copula density," *Journal of Multivariate Analysis*, vol. 101, pp. 200–222, Jan. 2010.

[40] L. Qu, Y. Qian, and H. Xie, "Copula density estimation by total variation penalized likelihood," *Communications in Statistics - Simulation and Computation*, vol. 38, no. 9, pp. 1891–1908, 2009.

[41] L. Qu and W. Yin, "Copula density estimation by total variation penalized likelihood with linear equality constraints," CAAM, Rice University, Tech. Rep. TR09-40, 2009. [Online]. Available: http://www.caam.rice.edu/~wy1/paperfiles/Rice_CAAM_TR09-40.PDF

[42] T. Chan, A. Marquina, and P. Mulet, "High-order total variation-based image restoration," *SIAM Journal on Scientific Computing*, vol. 22, no. 2, pp. 503–516, 2000.

[43] P. L. Combettes and J. C. Pesquet, "Image restoration subject to a total variation constraint," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1213–1222, Sep. 2004.

[44] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Trans. Image Process.*, vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[45] ——, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[46] ——, "Gradient-based algorithms with applications to signal recovery problems," in *Convex Optimization in Signal Processing and Communications*, D. P. Palomar and Y. C. Eldar, Eds. Cambridge university press, 2011, pp. 42–88.

[47] J. Nocedal and S. Wright, *Numerical Optimization*, 2nd ed. New York: Springer, 2006.

[48] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, 2009.

[49] J. J. Moreau, "Proximitet dualit dans un espace hilbertien," *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.

[50] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, pp. 1168–1200, 2005.

[51] A. Chambolle, "An algorithm for total variation minimization and applications," *Journal of Mathematical Imaging and Vision*, vol. 20, no. 1-2, pp. 89–97, 2004, special issue on mathematics and image analysis.

[52] R. Courant, "Variational methods for the solution of problems with equilibrium and vibration," *Bull. Amer. Math. Soc.*, vol. 49, pp. 1–23, 1943.

[53] Y. Wang, J. Yang, W. Yin, and Y. Zhang, "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, pp. 248–272, 2008.

[54] Z. Botev, J. Grotowski, and D. P. Kroese, "Kernel density estimation via diffusion," *Annals of Statistics*, vol. 38, pp. 2916–2957, 2010.

[55] Z. Botev, 2011, http://www.mathworks.com/matlabcentral/fileexchange/17204-kernel-density-estimation.