Reconstructing Paper Machine Sheet Process Data Variation Using Compressive Sensing

Parisa Towfighi∗ Guy Dumont ∗∗ Michael Davies Bhushan Gopaluni Philip Loewen

∗ University of British Columbia, Vancouver, BC, Canada, (Tel: 604-224-2616; e-mail: parisat@ece.ubc.ca).
∗∗ University of British Columbia, Vancouver, BC, Canada, (Tel: 604-822-8564; e-mail: guyd@ece.ubc.ca).

Abstract: During paper manufacture, system actuators need to control the properties of the entire sheet based on a restricted set of data measured by a scanning sensor that traverses the moving sheet. Cross direction variations (CD) are those along an axis perpendicular to the motion of the sheet, while machine direction (MD) variations are those along the axis of motion, and are assumed uniform in CD. Current industrial practice is to separate the relatively slow variations of the CD profile from the higher bandwidth MD variations using low pass filtering, although the spacing and timing of the scanned data measurements makes it inevitable that some process variations will be distorted or lost to aliasing in the filtered data. In this paper, a novel approach to estimation of MD and CD variations is proposed - compressive sensing. In this approach, knowledge of the process is used to help characterize the expected process variations, allowing accurate reconstruction of the true process variations from far fewer measurements than would be indicated by simple bandwidth-based uniform sampling theory. Instead, a random sampling protocol is used to accurately reconstruct the sheet properties. The approach is found to be effective, using simulated and actual industrial process data.

1. INTRODUCTION

Modern paper machines cannot be operated properly without rigorous control of the paper quality. A fast newsprint machine can produce a 10m wide roll of paper at a speed of 120 km/h. The characteristics of the end paper are described in two dimensions, the cross-direction (CD), across the width of the sheet and the machine direction (MD), in the direction that the sheet moves (Brewster et al. [1993]). A sketch of a simplified paper machine is shown (Fig. 1).

Fig. 1. Simplified sketch of a paper machine, (Ghofraniha [1997])

Paper properties are measured by a scanning sensor at up to 2000 locations across the sheet (Mijanovic et al. [2003]). Since the paper is moving forward while the sensor is traversing the width of the sheet, measurements are gathered along a zig-zag path over the sheet (Fig. 2) at a very shallow angle (typically less than 1 deg) with the machine direction. The main properties measured and to be controlled are basis weight, moisture and caliper (Brewster et al. [1993]). As they are taken along the above-mentioned path, the sensor measurements first have to be mapped into both CD and MD directions (Featherstone et al. [2000]) before being used by CD and MD controllers that manipulate either equally spaced actuators distributed in the CD direction for CD control, or MD actuators such as a stock valve for weight or dryer steam pressure for moisture.

Fig. 2. Scanner traverses the rolling sheet and gathers data in a zigzag path (Aslani et al. [2009])

Underlying the decomposition of the scanner measurements into orthogonal CD and MD variations is the assumption that the CD profile is nearly time invariant and the MD component is the same for all CD locations. Current industrial practice is to use exponential low pass filtering along the MD for each CD location (Featherstone et al.
1.1 An overview of compressive sensing

The Shannon theorem states that the sampling rate must be at least twice the highest frequency component of the signal. Twice the highest frequency component is known as the Nyquist rate. Compressive sensing theory claims that if we know about some general properties of the signal such as sparsity, we can reconstruct the signal from fewer measurements and have a reconstruction better than what was promised by the Shannon theorem. To use compressive sensing successfully, we need to meet two fundamental conditions: sparsity and incoherent sampling (Candes and Romberg [2007]). Compressive sensing exploits the property that many signals are sparse or compressible in a certain basis. Consider a signal \( s \) in the form of an \( n \times 1 \) column vector:

\[
\mathbf{s} = \mathbf{Ψ} \mathbf{x}.
\]

Here, \( \mathbf{x} \) is the \( n \times 1 \) column vector of weighing coefficients and the \( n \) columns of matrix \( \mathbf{Ψ} \) are orthonormal. In more detail,

\[
\mathbf{s} = \sum_{i=1}^{n} x_i \mathbf{ψ}_i.
\]

Here \( x \) is the coefficient sequence of \( s \), given by

\[
x_i = \langle s, \mathbf{ψ}_i \rangle
\]

Our signal \( s \) needs to be sparse in some orthonormal representation basis \( \mathbf{ψ} \), such as Fourier or wavelets. The signal \( s \) is \( K \)-sparse if it has \( K \) nonzero coefficients out of \( n \) coefficients. It is compressible if the representation has just a few large coefficients and many small coefficients.

The implication of sparsity is that we can ignore the small coefficients and still be able to usefully reconstruct the signal. Since we do not know the position of the sparse coefficients, it can be shown that a sufficient condition for recovery is the restricted isometry property.

For each integer \( K = 1, 2, \ldots \) define the isometry constant \( \delta_K \) of a matrix \( \mathbf{H} \) as the smallest number such that

\[
1 - \delta_K \leq \frac{\| \mathbf{H} \mathbf{x} \|^2}{\| \mathbf{x} \|^2} \leq 1 + \delta_K
\]

holds for all \( K \)-sparse vectors \( \mathbf{x} \). We say that a matrix \( \mathbf{H} \) has the restricted isometry property (RIP) of order \( K \) if \( \delta_K \) is not too close to 1 (usually less than 0.5). Generally, one method to obtaining a matrix \( \mathbf{H} \) that satisfies the RIP is to use a random projection matrix and obtain enough samples, see Candes [2008].

Depending on the sparsity of the signal in a given basis, the number of measurements that is likely to suffice for an acceptable reconstruction varies. However, in many practical situations, it is impossible to know the sparsity of a signal from just a few samples with certainty and as a result, it becomes difficult to know how many samples are needed. In theory, the number of measurements needed is \( M \), where

\[
M \gg cK \log \frac{n}{K}.
\]

Here, \( K \) is the sparsity level of the signal. The total number of points available in the original signal is denoted by \( n \), see Baraniuk [2007]. The oversampling factor is denoted by \( c \) and \( c > 1 \). For more information on how to select a proper oversampling factor, please refer to Baron et al. [2005]. Equation (5) implies that the sparser the signal, the fewer measurements are needed for successful reconstruction. In certain situations the sparsity rate varies and it becomes difficult to prove mathematically that RIP has been met.

Now let us consider incoherence. The coherence, \( μ \), is given by the equation below and it measures the maximum correlation between any two elements of \( ψ \) and \( φ \).

\[
μ(\Phi, \Psi) = \sqrt{n} \max_{k,j} |\langle \psi_k, \psi_j \rangle|.
\]

When \( \Phi \) and \( Ψ \) are orthonormal, the minimum coherence is 1 and the maximum coherence is \( \sqrt{n} \). Examples of pairs with maximal incoherence are spikes and sinusoids (Candès and Wakin [2008]). Consider a signal \( s \) being sampled by a projection matrix \( Φ \):

\[
y = \Phi \mathbf{s} + w.
\]

Here \( y \) is the measurement vector, \( w \) is the noise, and \( Φ \) is the projection matrix. Let us define a matrix \( H \) (Angelosante et al. [2009]) and call it a regression matrix:

\[
H = \Phi \Psi.
\]

Both the RIP and the incoherence condition can be met with high probability by selecting a random measurement matrix. If the measurement matrix contains independent and identically distributed random variables from a Gaussian probability density function with zero mean and variance 1/n, and if there are enough samples available, the regression matrix \( H \) satisfies RIP (Baraniuk [2007]).

\[
y = \Phi \mathbf{s} + w = \Phi \Psi \mathbf{x} + w = \mathbf{H} \mathbf{x} + w
\]

If we know the sampled data set \( y \), and the projection matrix \( Φ \), we can use a representation matrix \( Ψ \) to obtain \( \mathbf{H} \) and then recover the signal coefficients \( \mathbf{x} \). From the coefficients we can then reconstruct the original signal \( s \). This task can be accomplished via multiple algorithms. L^1 optimization is suitable for signals that are sparse. Let us consider the basis pursuit method.

Basis Pursuit (BP)

\[
\min \| x \|_1 \text{ s. t. } \mathbf{H} \mathbf{x} = \mathbf{y}
\]
where \( \| x \|_1 = \sum_{i=1}^{n} |x_i| \). This problem is convex and can be solved as a linear program. The basis pursuit method minimizes the coefficients \( x \) such that the condition \( Hx = y \) holds. It can be shown that the basis pursuit method is robust to measurement noise and discretization error as long as the oversampling rate, \( c \), is large enough (Baron et al. [2005]). For more information about the basis pursuit linear convex problem and the conditions under which useful reconstruction is feasible, please refer to Bajwa [2009].

There are solvers available which use recursive linear optimization algorithms (VanAntwerp and Braatz [2000]) to find the coefficients \( x \) given the regression matrix and the sampled data.

If there is a known level of noise in the signal, it is possible to have basis pursuit with denoising (BPDN). In this case a new parameter called \( \sigma \) is introduced that corresponds to the noise level.

\[
\min \| x \|_1 \quad \text{s. t.} \quad \| Hx - y \|_2 \leq \sigma \tag{11}
\]

BPDN is a quadratically constrained linear program. The same problem can be formulated differently when there is an estimate of how much noise is present in the coefficients of the original signal. In this case, one can include a tolerance parameter called \( \tau \) to minimize the coefficients while accounting for noise. This new method is called the Least Absolute Shrinkage and Selection Operator (LASSO), see Tibshirani [1996].

\[
\min \| y - Hx \|_2^2 \quad \text{s. t.} \quad \| x \|_1 \leq \tau \tag{12}
\]

LASSO is a quadratic program. Convex analysis can show that a solution to this problem is a minimizer of \( x \). The compressive sensing solver used in this paper is Spgl1 (van den Berg and Friedlander [2007]) employed in conjunction with Sparco (van den Berg et al. [2007]), which is a toolbox for solving sparse optimization problems. Both of these are MATLAB library tools and were developed at The University of British Columbia.

2. COMPRESSIVE SENSING AND PAPER SHEET ESTIMATION

When the scanning frequency is not at least twice the maximum frequency of the paper variation, it becomes difficult to estimate the sheet profile. To overcome this limitation, compressive sensing is used.

A sheet of paper is the original signal in a spatial domain. Two scenarios are tested. In one scenario a sheet of paper is generated based on available information regarding paper properties. In another case, a brown sheet of paper is scanned at all data points and its basis weight and moisture are recorded. The sheet of paper in both cases is represented in the form of a matrix. Samples are collected from the sheet either randomly or along the scanner path as shown in Fig. 2.

The MD and CD location of the sampled data is known and collected in a projection matrix. We define the regression matrix \( H \) to be the product of the measurement matrix in the spatial Dirac basis and another sparsifying basis such as Fourier or wavelet. The basis functions tested are Fourier, and three wavelet families: Daubechies, Haar and Symlet (Mallat [1999]).

The problem of reconstructing the paper properties from a small sample set can be formulated as minimizing the \( L^1 \) norm of the sparse coefficients, \( x \), such that the regression matrix, \( H \), applied to \( x \), yields the measured data vector \( y \).

\[
\min \| x \|_1 \quad \text{s. t.} \quad Hx = y \tag{13}
\]

With the regression matrix \( H \) and the measured data \( y \) being known, we can minimize the coefficients in \( x \). These coefficients represent the sheet of paper in a sparsifying basis such as Fourier or wavelet.

The regression matrix, projection matrix and the sparsifying basis are all passed to Sparco and wrapped as matrix operators. The regression matrix operator and the gathered samples are passed to the Spgl1 solver and the coefficients of the original signal are estimated. These coefficients are then transformed to obtain the paper sheet in a spatial domain. The error in compressive sensing method is calculated with all the tested basis functions.

To calculate the error inherent in the estimation, the original signal needs to be known. Samples taken from the original signal are used to estimate the entire signal. Then, the estimated signal is compared against the original signal. We calculate the relative error using the following formula.

\[
\text{error} = \frac{\| s - s_r \|}{\| s \|} \tag{14}
\]

Here, \( s \) is the original signal and \( s_r \) is the recovered signal. The compressive sensing results will be discussed shortly.

3. SIMULATED DATA

To test the compressive sensing method, a two dimensional paper properties signal is simulated. The simulated signal is a product of sinusoids; two sinusoids in the machine direction and one in the cross direction. The paper model was simulated based on information described in Farahmand et al. [2009].

The two dimensional signal is reshaped to be a row vector. Samples are collected at random and put in a column vector. The simulated paper sheet is shown (Fig. 3).

![Fig. 3. Original simulated paper sheet](image-url)
the true value generated. From the error plot (Fig. 4), one can conclude that not surprisingly increasing the number of measurements, results in a better approximation. Out of the different basis functions, Fourier yields the lowest error. This is to be expected, since the signal is mainly sinusoidal and therefore, most sparse in a Fourier basis.

Fig. 4. Error using compressive sensing with different sample sizes and basis functions

3.1 Overcoming Nyquist Rate by Compressive Sensing

As noted above, current industrial practice is to use a scanning sensor to gather sheet data and feed machine direction and cross direction variations to an exponential filter. The most common filter is a first order filter that separates out the lower frequency CD variation. The mean of each scan is the estimated MD value. When the sheet is being scanned in a zigzag manner, for measurement points that lie on the center line of the sheet, there is regular equal spacing between measurement points. All other measurement points have unequal but periodic spacing between consecutive measurements. This is because of the zigzag nature of the sensor’s path over the sheet.

To account for the varying distance between consecutive points, variable weighting coefficients are used. For more information about this adaptive exponential filter, refer to Aslani et al. [2009]. An alternative low pass filter with a weighting function is discussed in Duncan and Wellstead [2004].

In this section, compressive sensing is compared with exponential filtering. The paper variation is represented at 512 points in machine direction and 32 points in cross direction. Based on this sampling frequency two separate paper signals are simulated. Two scenarios are tested. In one scenario, the Nyquist limit cannot be met and in the other case, the Nyquist condition is satisfied. Nyquist limit cannot be met on the paper shown in Fig. 5 and Fig. 6.

Fig. 5. Simulated paper sheet when the Nyquist condition is not met

Fig. 6. Simulated paper sheet when the Nyquist condition is not met with noise, SNR=28.89

Fig. 7. Relative error when the Nyquist condition is met using exponential filtering and compressive sensing in the Fourier basis

methods, the lowest error is given by scanning a noiseless signal and reconstructing it using compressive sensing. Exponential filtering yields very high errors when the Nyquist rate is not met. Let us now consider a situation in which the Nyquist condition is satisfied. A plot of the original paper signal is shown in Fig. 8 and then from a set sample size, reconstruction is performed with exponential filtering and compressive sensing. A bar graph of the error is illustrated in Fig. 9. It can be observed that compressive sensing yields a better approximation than exponential
Fig. 8. Simulated paper sheet when the Nyquist condition is met.

Fig. 9. Relative error when the Nyquist condition is NOT met using exponential filtering and compressive sensing in the Fourier basis, SNR of noisy signal is 26.1222.

Fig. 10. Basis weight measurements are collected at all grid points of a sheet.

Fig. 11. Error calculated comparing the recovered and real basis weight data.

Fig. 12. Moisture measurements are collected at all grid points of a sheet.

Content of a roll after averaging the five machine direction scans. Fig. 11 and Fig. 13 show the relative reconstruction errors resulting from different basis functions. The basis functions are all assessed using random spatial sampling. Fourier basis is also tested with sampling by a simulated scanner along a zig-zag path. Sampling random points yields a better result as the measurement matrix can satisfy the RIP property with more probability. Since the variation in paper quality is expected to be smooth and of a sinusoidal nature, Fourier yields the best approximation. Also, the error decreases as sample size is increased.

4. INDUSTRIAL DATA

A large scale stationary section of a roll of unbleached paper was tested in an experiment carried out by Honeywell Inc. Detailed machine direction and cross direction measurements were gathered by a specially programmed scanning sensor.

There were five measurements gathered per point in the machine direction. The average of the five measurements was used. The sensors measured moisture and basis weight variations. This industrial data is depicted in Fig. 10 and Fig. 12 corresponding to basis weight and moisture content of a roll after averaging the five machine direction scans. Fig. 11 and Fig. 13 show the relative reconstruction errors resulting from different basis functions. The basis functions are all assessed using random spatial sampling. Fourier basis is also tested with sampling by a simulated scanner along a zig-zag path. Sampling random points yields a better result as the measurement matrix can satisfy the RIP property with more probability. Since the variation in paper quality is expected to be smooth and of a sinusoidal nature, Fourier yields the best approximation. Also, the error decreases as sample size is increased.

The compressive sensing needs to estimate paper properties as the paper moves through a paper machine. It takes at least 15-60 seconds for one scan. Computational time can be adjusted by choosing how often to run the solver, and also the size of the paper properties signal to be estimated.
The compressive sampling rate is useful in situations where the Nyquist sampling cannot be achieved, but conditions of sparsity and incoherence are likely to be met with high probability. The effectiveness of the compressive sensing method with paper sheet data was illustrated and it was shown that this method is more accurate than exponential filtering when applied to both simulated and industrial data. The data represented variation in paper weight and moisture.

The disadvantage of the approach used here is that due to the nature of the inverse problem, the solver recovers coefficients for square signals only and some signal interpolation and averaging may be necessary in some cases. Work is continuing on the implementation of compressive sensing on paper machine data in real time.

ACKNOWLEDGEMENTS

Many special thanks to Danlei Chu and Johan Backstrom from Honeywell Process Solutions for providing industrial data and to Dr. Felix J. Herrmann for lectures on compressive sensing.

REFERENCES