Optimal Signal Demixing

Applications for Energy Data

Bennet Meyers

Qualifying Examination, Electrical Engineering Dept., Stanford University

October 2019

Table of Contents

Table of Contents

Background: More Data, More Opportunities

- Increasing volume of photovoltaic (PV) system performance data creates opportunities for monitoring system health and optimizing operations and maintenance (O&M) activities.
- Digital O&M \$9b industry by 2024 ("The State of Digital O&M for the Solar Market", *Greentech Media*, 10/10/19)
- However, classic approaches—waterfall analysis, performance index analysis—require
  - A significant amount of engineering time
  - Knowledge of PV system modeling science and best practices
  - Accurate system configuration information
  - Access to reliable irradiance and meteorological data

New Approaches are Needed

For these reasons, existing PV system performance engineering methods are focused on utility scale power plants...

...rather than the rapidly increasing number of distributed rooftop systems.

Image credit: SunPower Corp.

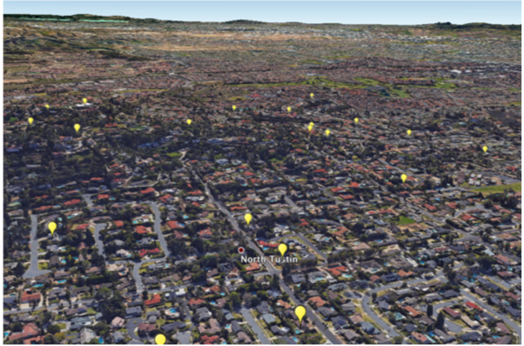


Image credit: Google Earth

Utility vs. Distributed

	Utility	Distributed
Site model	✓	✗
Irradiance data	✓	✗
Meteorological data	✓	✗
People / PV system	> 1	$\ll 1$

- New approaches needed to analyze and managed distributed PV
- How to extract insight into system health from only a power signal?

My Work So Far

- Proposed "PVInsight" to the U.S. Department of Energy Solar Energy Technologies Office (DOE SETO), which was successfully funded over two years (FY 2019-2020)
- One journal paper and one conference paper
  - B. Meyers, M. Deceglie, C. Deline, and D. Jordan, "Signal Processing on PV Time-Series Data: Robust Degradation Analysis without Physical Models," *IEEE Journal of Photovoltaics* (accepted)
  - B. Meyers, M. Tabone, and E. C. Kara, "Statistical Clear Sky Fitting Algorithm," *Proc. of IEEE 45th Photovolt. Spec. Conf.*, 2018.
- Three conference oral presentations
  - IEEE Photovoltaics Specialists Conference (2018, 2019)
  - NREL Reliability Workshop (2019)
- Two published open-source Python projects
  - `solar-data-tools`: `https://github.com/slacgismo/solar-data-tools`
  - `StatisticalClearSky`: `https://github.com/slacgismo/StatisticalClearSky`

Table of Contents

Background on Demixing Signals

$$y_t = (x_0)_t + \cdots + (x_K)_t \in \mathbf{R}^p$$

- Given y_t, estimate its components x_k
- Generally a highly underdetermined problem, but many useful approaches are possible
  - Filtering, PCA, ICA, NMF, basis pursuit
- Common themes:
  - Maximize independence or minimize correlation between components
  - Impose low-complexity constraints on components

Background on Demixing Signals

$$y_t = (x_0)_t + \cdots + (x_K)_t \in \mathbf{R}^p$$

- Given y_t, estimate its components x_k
- Generally a highly underdetermined problem, but many useful approaches are possible
  - Filtering, PCA, ICA, NMF, basis pursuit
- Common themes:
  - Maximize independence or minimize correlation between components
  - Impose low-complexity constraints on components

A Brief History

- Many names: blind source separation, disaggregation, decomposition, trend estimation...
- Fourier analysis formalized in late 18th to early 19th century (Fourier, Lagrange, Gauss...)
- PCA invented by Karl Pearson in 1901
- Trend-cycle decomposition and smoothing developed by Frederick Macaulay in the 1920s
- Sinc interpolation and low-pass filtering developed by Whittaker, Nyquist, Kotelnikov, Shannon in 1930s-1940s
- Colin Cherry defined and named the "cocktail party problem" in 1953
- ICA first developed by Jeanny Hérault and Bernard Ans in mid-1980s
    - Refined by Hyvärinen in 1999
- Basis pursuit proposed by Chen, Donoho, and Saunders in 2001
- Contextually supervised source separation proposed by Wytock and Kolter in 2013

Optimal Signal Demixing (OSD) Problem

$$\begin{array}{ll} \underset{\substack{x_0,\ldots,x_K \\ \theta_0,\ldots,\theta_K}}{\text{minimize}} & \sum_{k=0}^{K} \lambda_k \phi_k \left(x_k, \theta_k\right) \\ \text{subject to} & y_{i,t} = \sum_{k=0}^{K} \left(x_k\right)_{i,t} \quad \text{for} \quad (i,t) \in \mathcal{K} \end{array}$$

- Optimize over variables $x_k \in \mathbf{R}^{p \times T}$ and $\theta_k \in \mathbf{R}^{m_k}$ for $k = 0, \ldots, K$
  - x_k are demixed components with x_0 being the noise term
  - θ_k are optional extra variables for parameterized models
- $y \in \mathbf{R}^{p \times T}$ is the observed signal (problem data) with set of known entries $\mathcal{K}$
- $\phi_k : \mathbf{R}^{p \times T} \to \mathbf{R} \cup \{\infty\}$ defines the class of component x_k
- $\lambda_k \geq 0$ trade off between importance of components' costs ($\lambda_0 = 1$ by convention)
- If all ϕ_k are convex in x_k and θ_k, it is a convex optimization problem

Component Classes

- Smaller $\phi_k(x) \to$ more plausible x_k is
- $\lambda_k \phi_k(x)$ can be interpreted as negative log-likelihood (assuming independence)
- Constraints can be incorporated in any class cost function by defining the function to be equal to infinity if constaints are violated
- For many useful choices of ϕ_k, OSD is a quadratic program (QP)

Noise Component Classes ($p = 1$)

Class Description	$\phi(x)$
Small, Gaussian	$\sum_{t=1}^{T} x_t^2$
Small, Laplacian (“spikey”)	$\sum_{t=1}^{T} \lvert x_t \rvert$
Small, skewed, set quantile τ	$\sum_{t=1}^{T} \left[\frac{1}{2}\lvert x_t\rvert + \left(\tau - \frac{1}{2}\right) x_t\right]$
Huber (thick-tailed Gaussian), set breakpoint a	$\sum_{t=1}^{T} \left[\begin{cases} \frac{1}{2} x_t^2 & \lvert x_t\rvert \leq a \\ a\left(\lvert x_t\rvert - \frac{a}{2}\right) & \lvert x_t\rvert > a \end{cases}\right]$

Structured Component Classes—Convex Examples ($p = 1$)

Class Description	$\phi(x)$
Smooth, slowly changing average	$\sum_{t=1}^{T-1} (x_{t+1} - x_t)^2$
Smooth, slowly changing slope	$\sum_{t=1}^{T-2} (x_{t+2} - 2x_{t+1} + x_t)^2$
Piecewise constant (heuristic)	$\sum_{t=1}^{T-1} \lvert x_{t+1} - x_t \rvert$
Piecewise linear (heuristic)	$\sum_{t=1}^{T-2} \lvert x_{t+2} - 2x_{t+1} + x_t \rvert$
Periodic, period q	$x_{t+q} = x_t$ for $t = 1, \ldots, T - q$
Monotonically increasing	$x_{t+1} \geq x_t$ for $t = 1, \ldots, T - 1$
Autoregressive, known coefficients α_i	$x_t = \alpha_1 x_{t-1} + \cdots + \alpha_r x_{t-r}$ for $t = r + 1, \ldots, T$
Linear model, new parameter θ	$x = A\theta$

Structured Component Classes—Non-Convex Examples ($p = 1$)

Class Description	$\phi(x)$
Piecewise constant (exact)	$\sum_{t=1}^{T-1} \mathbf{1}\left(x_{t+1} - x_t \neq 0\right)$
Piecewise linear (exact)	$\sum_{t=1}^{T-2} \mathbf{1}\left(x_{t+2} - 2x_{t+1} + x_t \neq 0\right)$
Autoregressive, unknown coefficients θ_i	$x_t = \theta_1 x_{t-1} + \cdots + \theta_r x_{t-r}$ for $t = r+1, \ldots, T$
Discrete feasible values in set $\mathcal{F}$	$x_t \in \mathcal{F}$ for $t = 1, \ldots, T$

Classic Examples of Filtering and Smoothing ($K = 1$)

- Hodrick-Prescott filtering[1]
  - $\phi_0(x) = \sum_{t=1}^{T} x_t^2$
  - $\phi_1(x) = \sum_{t=1}^{T-2} (x_{t+2} - 2x_{t+1} + x_t)^2$
- Total variation denoising[2]
  - $\phi_0(x) = \sum_{t=1}^{T} x_t^2$
  - $\phi_1(x) = \sum_{t=1}^{T-1} |x_{t+1} - x_t|$
- Quantile smoothing[3] (set quantile $0 < \tau < 1$)
  - $\phi_0(x) = \sum_{t=1}^{T} \left[\frac{1}{2} |x_t| + \left(\tau - \frac{1}{2}\right) x_t\right]$
  - $\phi_1(x) = \sum_{t=1}^{T-2} (x_{t+2} - 2x_{t+1} + x_t)^2$

---

[1] C. E. V. Leser, "A Simple Method of Trend Construction," 1961.

[2] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," 1992.

[3] K. Abberger, "Quantile smoothing in financial time series," 1997.

Missing Data: Imputation and Validation

- Problem data $y \in \mathbf{R}^{p \times T}$ may have missing entries (?, NaN, etc.)
- $(i, t) \in \mathcal{K}$ are the known entries of y
- Imputation:
  - Solve OSD problem to estimate $x_0, \ldots, x_K$
  - For $(i, t) \notin \mathcal{K}$, we guess $\hat{y}_{i,t} = (x_0)_{i,t} + \cdots + (x_K)_{i,t}$
- Validation
  - Hold out some entries of y to make set $\mathcal{K}'$, and then compare with imputed value
  - Measure: $\sum_{(i,t)\in\mathcal{K}'} \phi_0(y_{i,t} - \hat{y}_{i,t})$

Convex Heuristics for Non-convex Cases

- Many useful ϕ_k are convex
- We can evaluate non-convex ϕ_k, but typically can't solve exactly
- ℓ_1-norm approximation for cardinality minimization problems[4]
  - All total-variation type problems: edge detection, piecewise constant functions, piecewise linear functions, lasso regression
  - Compatible with polishing, iterative reweighting, and other heuristics for improving sparsity
- ADMM for mixed-integer constraints[5]
- Alternating convex optimization for biconvex ϕ_k's

[4] Candès, et al., "Enhancing sparsity by reweighted L1 minimization," 2008.

[5] Takapoui, et al., "A simple effective heuristic for embedded mixed-integer quadratic programming," 2017.

Convex Heuristics for Non-convex Cases

- Many useful ϕ_k are convex
- We can evaluate non-convex ϕ_k, but typically can't solve exactly
- ℓ_1-norm approximation for cardinality minimization problems[4]
  - All total-variation type problems: edge detection, piecewise constant functions, piecewise linear functions, lasso regression
  - Compatible with polishing, iterative reweighting, and other heuristics for improving sparsity
- ADMM for mixed-integer constraints[5]
- Alternating convex optimization for biconvex ϕ_k's

[4] Candès, et al., "Enhancing sparsity by reweighted L1 minimization," 2008.

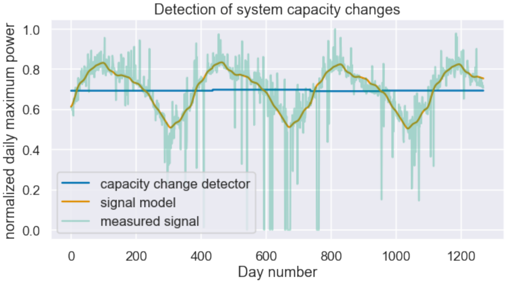
[5] Takapoui, et al., "A simple effective heuristic for embedded mixed-integer quadratic programming," 2017.

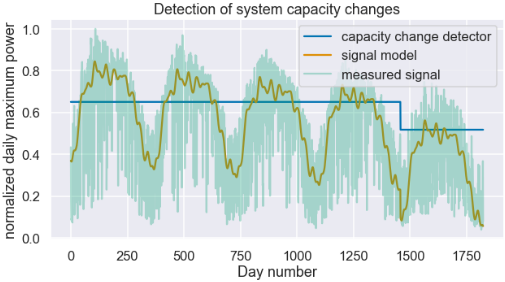
Table of Contents

System Capacity Change Detection

Problem: Detect equipment failure as a step-change in apparent PV system capacity

Component	*Class*	$\phi(x, \theta)$	λ
x_0	small, Laplacian	$\sum_{t=1}^{T} \lvert x_t \rvert$	1
x_1	smooth and 365-day periodic with linear offset	$\begin{cases} \sum_{t=1}^{T-2} (x_{t+2} - 2x_{t+1} + x_t)^2 + \theta^2 & x_{t+365} - x_t = \theta \\ \infty & x_{t+365} - x_t \neq \theta \end{cases}$	15
x_2	iteratively weighted ($w_t^{(j)}$) piecewise constant	$\sum_{t=1}^{T-1} \left\lvert w_t^{(j)} (x_{t+1} - x_t) \right\rvert$	100

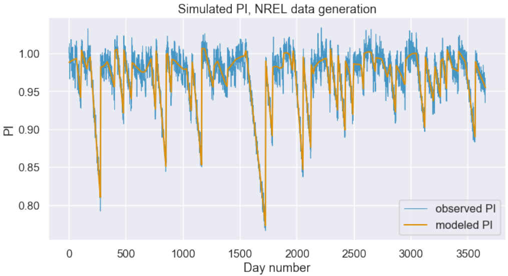




Soiling Analysis

Problem: Estimate soiling rates and energy loss from system performance index (PI)

Component	*Class*	$\phi(x,\theta)$	λ
x_0	small, Gaussian	$\sum_{t=1}^{T} x_t^2$	1
x_1	smooth and 365-day periodic	$\begin{cases} \sum_{t=1}^{T-2}(x_{t+2}-2x_{t+1}+x_t)^2 & x_{t+365}=x_t \\ \infty & x_{t+365}\neq x_t \end{cases}$	10
x_2	iteratively weighted $(w_t^{(j)})$ piecewise linear	$\sum_{t=1}^{T-2}\left\lvert w_t^{(j)}(x_{t+2}-2x_{t+1}+x_t)\right\rvert$	0.1



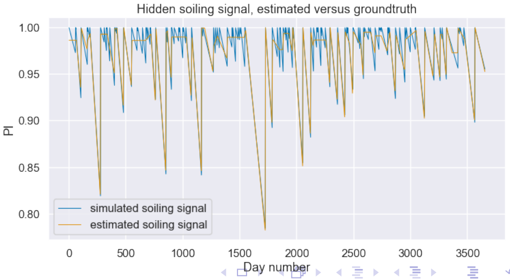


Table of Contents

Next Steps

- Continued literature review to establish connections with previous work
- Manuscript on proposed framework
- Python software package implementing framework, to be referenced in manuscript
- Continued application of methods to energy analysis problems
- Apply to forecasting and MPC problems

Appendix A: Additional background and papers on signal demixing

Background on Signal Separation

- The separation of a set of components from a set of mixed signals
- Colin Cherry defined and named the "cocktail party problem" in 1953[6]
- Generally a highly underdetermined problem, but many useful approaches are possible
- Problems are generally categorized by
  - The number of observed channels (single-channel or multi-channel)
  - The mixture process (linear instantaneous, delayed, convolutive, or non-linear)
- Common themes
  - Look for components that are maximally independent or minimally correlated
    - Principle component analysis, independent component analysis
  - Impose some kind of low-complexity constraint on components
    - Non-negative matrix factorization, basis pursuit, filtering methods

[6] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *J. Acoust. Soc. Am.*, vol. 25, no. 5, pp. 975–979, Sep. 1953.

Background: Independent Component Analysis

- First proposed by Jeanny Hérault and Bernard Ans in around 1982–1984 (in French)
- Refined and clarified by Pierre Comon in 1994[7]
- Fast and efficient algorithm proposed by Anthony Bell and Terry Sejnowski in 1995[8]
- Modern interpretation is based on maximum likelihood estimation, in which we assume that sources are independent and non-Gaussian (typically logistic)
- Classically, ICA solves multi-channel, linear instantaneous problems, but other extensions have been proposed
- More recently single channel ICA has been explored, with mixed results[9]

[7] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[8] A. J. Bell and T. J. Sejnowski, "An Information-Maximization Approach to Blind Separation and Blind Deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[9] M. E. Davies and C. J. James, "Source separation using single channel ICA," *Signal Processing*, vol. 87, no. 8, pp. 1819–1832, 2007.

Background: Dictionaries and Basis Pursuit

- Building upon classic representation of periodic signals as collection of sinusoids, many alternate dictionaries of parameterized waveforms have been proposed
  - Wavelets, Gabor dictionaries, Cosine Packets, Chirplets, and many more
- 1988: Ingrid Daubechies proposed the Method of Frames for finding decompositions of signals into both time and frequency[10]
- 2001: Chen, Donoho, and Saunders proposed Basis Pursuit for decomposing a signal into an optimal superposition of dictionary elements[11]
- Chen et al. used convex optimization to find sparse signal representation, even with overcomplete dictionaries

[10] I. Daubechies, "Time-frequency localization operators: a geometric phase space approach," *IEEE Trans. Inf. Theory*, vol. 34, no. 4, pp. 605–612, Jul. 1988.

[11] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic Decomposition by Basis Pursuit," SIAM Rev., vol. 43, no. 1, pp. 129–159, Jan. 2001.

Appendix B: Additional examples

Classic Examples of Fitting and Smoothing ($k = 1$)

- Hodrick-Prescott filtering
  - $\phi_0(x_0) = \|x_0\|_2^2$
  - $\phi_1(x_1) = \left\|\Delta^2 x_1\right\|_2^2$
- Quantile smoothing
  - $\phi_0(x_0) = \tau \mathbf{1}^T (x_0)_+ + (\mathbf{1} - \tau)^T (x_0)_-$
  - $\phi_1(x_1) = \left\|\Delta^2 x_1\right\|_2^2$
- ℓ_1 trend filtering
  - $\phi_0(x_0) = \|x_0\|_2^2$
  - $\phi_1(x_1) = \left\|\Delta^2 x_1\right\|_1$
- Total variation filtering
  - $\phi_0(x_0) = \|x_0\|_2^2$
  - $\phi_1(x_1) = \left\|\Delta x_1\right\|_1$
- Autoregressive model, order r
  - $\phi_0(x_0) = \|x_0\|_2^2$
  - $\phi_1(x_1, \theta) = \begin{cases} 0 & (x_1)_t = \sum_{j=1}^r \theta_j (x_1)_{t-j} + \theta_0 \,\forall\, t \\ \infty & \textit{otherwise} \end{cases}$
- Linear regression to time
  - $\phi_0(x_0) = \|x_0\|_2^2$
  - $\phi_1(x_1, \theta) = \begin{cases} 0 & (x_1)_t = \theta_1 t + \theta_0 \,\forall\, t \\ \infty & \textit{otherwise} \end{cases}$
- Robust regression (B&V[12] example 6.2)
  - $\phi_0(x_0) = \sum_{t=1}^n \operatorname{huber}((x_0)_t)$
  - $\phi_1(x_1, \theta) = \begin{cases} 0 & (x_1)_t = \theta_1 t + \theta_0 \,\forall\, t \\ \infty & \textit{otherwise} \end{cases}$

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge, U.K.: Cambridge Univ. Press, 2004