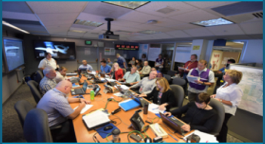
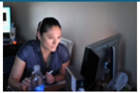
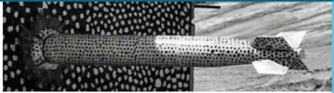
 10113C

Sandia National Laboratories

Using Filter Methods to Guide Convergence for ADMM, with Applications to Nonnegative Matrix Factorization Problems







Robert Baraldi, Sven Leyffer, Stefan Wild

ICCOPT 2022 - Nonlinear Optimization

Bethlehem, PA

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell

dia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly ow subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2022–xxxx C

Outline

General Nonlinear Optimization I

- Nonlinear Programs (NLP):

$$\boxed{\begin{aligned} \text{minimize}_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{such that} \quad & c(\boldsymbol{x}) = 0 \\ & \boldsymbol{x} \in \mathcal{X} \end{aligned}} \tag{1}$$

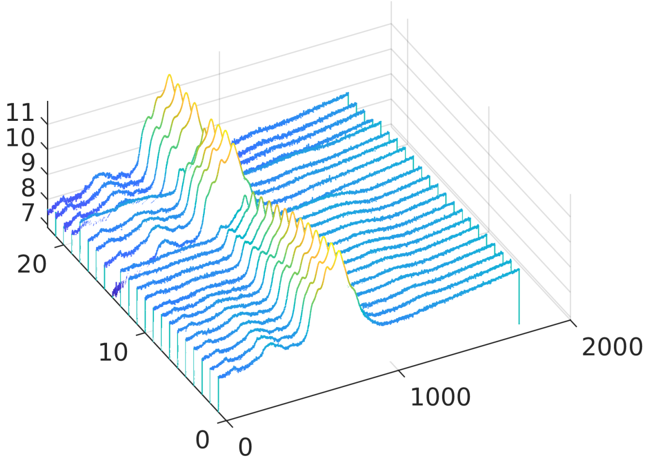
- $\boldsymbol{x} := [x_0; \ldots; x_p] \in \mathbb{R}^n$ for p blocks $(n = \sum_{i=0}^{p} n_i)$, $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$, $\mathcal{X}$ simple bounds.
- We assume $f, c \in \mathcal{C}^2$; f, c possibly separable.

Motivating Example - Chemical Spectrum Analysis

- Find distributions of chemicals that occur in measured spectrum analysis.
- Each chemical follows a Gaussian distribution; need nonnegative combination of Gaussians that reproduce the spectra data.
- *Intensity Function* as a function of wave number w and concentration c, with rank constraint K. For m be the number of concentrations c:

$$\hat{I}(w_i, c_j) = \sum_{k=1}^{K} \underbrace{Y_{k,j}}_{f_k(c_j)} \underbrace{\exp\left(-\frac{1}{2}\left(\frac{w_i - \mu_k}{\sigma_k}\right)^2\right)}_{g_k(w_i)} \tag{2}$$

- Built "image" from mean, standard deviations, and nonnegative coefficients



Spectra data: $M \in \mathbb{R}^{1750\times 22}$, concentration (x-axis) $\times$ wavelength (y-axis) $\times$ $\hat{I}$ (z).

Splitting-ADMM History/Applications

- First utilized in Douglas and Rachford [1956], extended as method for variational problems in Gabay and Mercier [1975], Glowinski and Marroco [1975].
- Classical ADMM takes the form of a 2-block update

$$\boxed{\min_{\boldsymbol{x}:=[x_1,x_2]} f(\boldsymbol{x}) := f_1(x_1) + f_2(x_2) \quad \text{subject to} \quad c(\boldsymbol{x}) := A_1x_1 + A_2x_2 - b = 0.} \tag{3}$$

- Most literature assumes $c(\boldsymbol{x}) = \mathbf{A}\boldsymbol{x} - b$ or linear constraints, and $f = f_1 + f_2$.

Direct multiblock extension

- Now $\boldsymbol{x} = [x_1, \ldots, x_p]$,

$$\boxed{\min_{x \in \mathcal{X}} f(\boldsymbol{x}) \quad \text{s.t. } A_0 x_0 + \ldots + A_p x_p = \mathbf{b}.} \tag{4}$$

simple constraint sets $\mathcal{X}$ (e.g., bounds), multiaffine $c(\boldsymbol{x})$ as in Gao et al. [2020].

- Block-minimize he augmented Lagrangian for (4)

$$\boxed{\mathcal{L}_\rho(\boldsymbol{x}, y) = f(\boldsymbol{x}) + y^T(\mathbf{A}\boldsymbol{x} - \mathbf{b}) + \frac{\rho}{2}\|\mathbf{A}\boldsymbol{x} - \mathbf{b}\|^2} \tag{5}$$

- The classical 2-block problem differs greatly; Chen et al. [2016] provides example $p = 3$ blocks $f_1, f_2, f_3 \equiv 0$ where ADMM diverges for any ρ.
- Nonconvex settings Wang et al. [2018]: NMF/C, phase retrieval, concensus problems.

P-block algorithm

Algorithm 1: Basic ADMM P-Block splitting for (4)

Data: functions f, matrices $\mathbf{A}$, vector $\mathbf{b}$, augmented Lagrangian parameter $\rho > 0$, and initial $\boldsymbol{x}_0, y_0$

Result: $\boldsymbol{x} \leftarrow \arg\min_{\boldsymbol{x}} f(\boldsymbol{x})$ s.t. $\mathbf{A}\boldsymbol{x} = \mathbf{b}$ in (4)

for $k = 0, 1, \ldots$ **do**

  for $i = 1, \ldots, p$ **do**

    $x_i^{(k+1)} = \arg\min_{x_i} \mathcal{L}_\rho([\boldsymbol{x}_{<i}^{(k)}, x_i^{(k)}, \boldsymbol{x}_{>i}^{(k)}], y^{(k)})$

  end

  $y^{(k+1)} = y^{(k)} + \rho(\mathbf{A}\boldsymbol{x}^{(k+1)} - \mathbf{b})$

end

Lecture Goals

- Literature: ADMM convergence is still “somewhat” nebulous: requires *a priori* knowledge of functions, ρ choice is largely open, limited choices in $c(\boldsymbol{x})$.
- Tools:
  - Augmented Lagrangian, ADMM/block-updates, Filter Methods

Problem: can we solve nonconvex, nonlinear problems (1) with block structure?

- What about nonconvex/nonlinear constraints?
- Can filters help guide convergence?

Basic Lagrangian Methods I

- Lagrangian of NLP (1)

$$\mathcal{L}(\boldsymbol{x}, y) = f(\boldsymbol{x}) + y^T c(\boldsymbol{x}) \tag{6}$$

  where $y \in \mathbb{R}^m$ is the vector of Lagrange multipliers of $c(\boldsymbol{x})$.

- $\mathcal{X} \subseteq \mathbb{R}^n$ is a simple feasible set:

$$\operatorname*{proj}_{\mathcal{X}}(\boldsymbol{x}) \equiv \operatorname*{arg\,min}_{y \in \mathcal{X}} \|y - \boldsymbol{x}\| \tag{7}$$

- First-order optimality & feasibility (w/ $\mathcal{X}$ simple bounds $l \leq \boldsymbol{x} \leq u$):

$$\operatorname*{proj}_{\mathcal{X}}(\boldsymbol{x} - \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}, y)) - \boldsymbol{x} = 0 \quad \Rightarrow \quad \min\{\boldsymbol{x} - l, \max\{\boldsymbol{x} - u, \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}, y)\}\} = 0 \tag{8a}$$

$$c(\boldsymbol{x}) = 0 \tag{8b}$$

Augmented Lagrangian Methods

- Adds quadratic penalty term to (6):

$$\mathcal{L}_\rho(\boldsymbol{x}, y) = \underbrace{f(\boldsymbol{x}) - y^T c(\boldsymbol{x})}_{\mathcal{L}_0(\boldsymbol{x},y)} + \frac{1}{2}\rho\|c(\boldsymbol{x})\|^2 \tag{9}$$

  for a penalty parameter ρ.
- Bound Constrained Aug-Lag:

$$\min_{\boldsymbol{x}} \quad \mathcal{L}_{\rho_k}(\boldsymbol{x}, y^{(k)}) \tag{10a}$$

$$\text{such that} \quad l \leq \boldsymbol{x} \leq u \tag{10b}$$

  for fixed ρ_k and $y^{(k)}$.
- Solve (10), and then

$$y^{(k+1)} = y^{(k)} - \rho c(\boldsymbol{x}^{(k+1)})$$

- $y^{(k)}$ update makes dual feasible, primal feasibility in limit

Augmented Lagrangian Formulations

- Backbone for solving nonlinear problems Vanaret and Leyffer [2020]:
  - Interior-point methods; most used for large-scale problems
  - Active-set methods; great for warm-starts
  - Issues: lack warm-starts (IP) and cannot scale (AS)

> Goal: Leverage different solver flexibilities for nonlinearities in f and c

- *But*: need scaling capabilities for independent block updates - destroyed by quadratic term

Filter Methods

- Monitoring trade-off between primal feasibility and first-order convergence.
- η and ω represent infeasibility and first order error:

$$\eta(x) := \|c(\boldsymbol{x})\| \tag{11a}$$

$$\omega(x, y) := \|\operatorname{proj}_{\mathcal{X}}(\boldsymbol{x} - \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}, y)) - \boldsymbol{x}\| \tag{11b}$$

- Keep track of (ω_0, η) pairs, update $y^{(k)}$ when reach acceptable point or ρ_k if not feasible enough

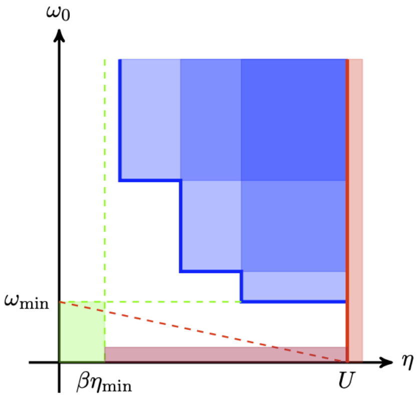
Definition (Augmented Lagrangian Filter and Acceptance)

A filter $\mathcal{F}$ is a list of pairs $(\eta_l, \omega_l) := (\eta(\boldsymbol{x}^{(l)}), \omega_0(\boldsymbol{x}^{(l)}, y^{(l)}))$ such that no pair dominates another pair, i.e., there exists no pairs $(\eta_l, \omega_l), (\eta_k, \omega_k)$, $l \neq k$ such that $\eta_l \leq \eta_k$ and $\omega_l \leq \omega_k$. A point $(\boldsymbol{x}^{(k)}, y^{(k)})$ is acceptable to the filter $\mathcal{F}$ iff

$$\eta_k := \eta(\boldsymbol{x}^{(k)}) \leq \beta\eta_l \text{ or } \omega_k := \omega_0(\boldsymbol{x}^{(k)}, y^{(k)}) \leq \omega_l - \gamma\eta(\boldsymbol{x}^{(k)}) \tag{12}$$

$\forall\ (\eta_l, \omega_l) \in \mathcal{F}$ where $0 < \gamma, \beta < 1$.

Filter Example



(a) Filter is blue; U is upper bound ($\eta(\boldsymbol{x}) \leq U := \max(\omega_{\min}/\gamma, \beta\eta_{\min})$). Values in the blue are not acceptable; shaded purple triggers restoration.

Filter Methods: Algorithm

- Monitors dual infeasibility error of the original problem while solving the augmented Lagrangian.
- Restoration conditions

$$\eta(\boldsymbol{x}) = \eta(\boldsymbol{x}^{(j+1)}) \geq \beta U \tag{13a}$$

$$\omega_{\rho_k}(\boldsymbol{x}^{(j+1)}, y^{(k)}) \leq \epsilon \quad \text{and} \quad \eta(\boldsymbol{x}^{(j+1)}) \geq \beta\eta_{\text{min}}. \tag{13b}$$

- *Optimal* stopping criteria: relative first order error ω_{ρ_k} (11b)

$$\frac{\| \operatorname{proj}_{\mathcal{X}}(\boldsymbol{x} - \nabla_{\boldsymbol{x}}\mathcal{L}_0(\boldsymbol{x}^{(k)}, y^{(k)})) - \boldsymbol{x}\|}{\| \operatorname{proj}_{\mathcal{X}}(\boldsymbol{x} - \nabla_{\boldsymbol{x}}\mathcal{L}_0(\boldsymbol{x}^{(0)}, y^{(0)})) - \boldsymbol{x}\|} = \frac{\omega_0(\boldsymbol{x}_k, y_k)}{\omega_0(\boldsymbol{x}_0, y_0)} \leq \varepsilon \tag{14}$$

where $\boldsymbol{x}^{(0)}$ is the initial iterate.

ADMM Filter Algorithm

Algorithm 2: Outline of ADMM-F for (1)

while $(\mathbf{x}^{(k)}, y^{(k)})$ *not optimal* (14) **do**
  while $(\eta^{(j)}, \omega^{(j)})$ *not acceptable to* $\mathcal{F}_k$ **do**
    while *! Satisfied* **do**
      for $i = 1, \ldots, p$ *with p blocks* **do**
        if $j == 0$ **then**
          $x_i^{(j+1)} \leftarrow \mathrm{proj}_{\mathcal{X}_i}(x_i^{(j)} - \alpha_i \nabla_i \mathcal{L}_{\rho_k}([\mathbf{x}_{<i}^{(j+1)}, x_i^{(j)}, \mathbf{x}_{>i}^{(j)}], y^{(k)}))$ `// Block coordinate steepest descent`
        else
          $x_i^{(j+1)} \leftarrow \arg\min_{x_i \in \mathcal{X}_i} \mathcal{L}_{\rho_k}([\mathbf{x}_{<i}^{(j+1)}, x_i, \mathbf{x}_{>i}^{(j)}], y^{(k)})$
      end
    end
    if *Restoration condition* (13) *holds* **then**
      Find $\mathbf{x}^{(j+1)}$ s.t. $(\eta^{(j+1)}, \omega^{(j+1)}) \in \mathcal{F}_k$, or $\mathbf{x}^{(j+1)} \leftarrow \arg\min_{x \in \mathcal{X}} \|c(\mathbf{x})\|^2$;
      `// exits Filter/j-loop`
    else
      Compute $\omega_{\rho_k}(\mathbf{x}^{(j)}, y^{(k)}), \eta(\mathbf{x}^{(j)}, y^{(k)})$;
  end
  if $\eta^{(k)} > 0$ **then**
    Add $(\eta^{(k)}, \omega^{(k)})$ to $\mathcal{F}_k$
    Update: $(\eta^{(k)}, \omega^{(k)}) \leftarrow (\omega_{\rho_k}(\mathbf{x}^{(k+1)}, y^{(k+1)}), \eta(\mathbf{x}^{(k+1)}, y^{(k+1)}))$;
  if *Entered Restoration* **then**
    Increase Penalty $\rho_{k+1} \leftarrow \max(\rho_k, \xi\rho_k)$;
end

Assumption (Differentiability & Set compactness)

Consider (1), *and assume that* f *and* c *are continuously differentiable. In addition, the constraint norm satisfies* $\|c(\boldsymbol{x})\| \to \infty$ *as* $\|\boldsymbol{x}\| \to \infty$.

Assumption 1 implies that our iterates remain in a compact set, replaced with finite bounds $l \leq \boldsymbol{x} \leq u$. Assumption 1 implies that f, c, and their derivatives are bounded for all iterates.
Algorithm 2 has three distinct outcomes:

1. There exists an infinite sequence of restoration phase iterates $\boldsymbol{x}^{k_l}$, indexed by $\mathcal{R} := \{k_1, k_2, \ldots\}$, whose limit point $\boldsymbol{x}^* := \lim \boldsymbol{x}^{k_l}$ minimizes the constraint violation, satisfying $\eta(\boldsymbol{x}^*) > 0$.
2. There exists an infinite sequence of successful major iterates $\boldsymbol{x}^{k_l}$, indexed by $\mathcal{S} := \{k_1, k_2, \ldots\}$, and the linear independence constraint qualification (LICQ) fails to hold at the limit $\boldsymbol{x}^* := \lim \boldsymbol{x}^{k_l}$, which is a Fritz-John (FJ) point of (1).
3. There exists an infinite sequence of successful major iterates $\boldsymbol{x}^{k_l}$, indexed by $\mathcal{S} := \{k_l, k_2, \ldots\}$ and LICQ holds at the limit $\boldsymbol{x}^* := \lim \boldsymbol{x}^{k_l}$, which is a KKT point of (1).

Convergence of Algorithm 2

Theorem

Under Assumption 1, either Algorithm 2 terminates afer a finite number of iterations at a KKT point: for some k, $\mathbf{x}^{(k)}$ is a first-order stationary point with $\eta(\mathbf{x}^{(k)}) = 0$ and $\omega(\mathbf{x}^{(k)}) = 0$, or there exists an infinite seuqnce of iterates $\mathbf{x}^{(k)}$ and any limit point $\mathbf{x}^{(k)} \to \mathbf{x}^*$ that satisfy one of the following:

1. *The penalty parameter is updated finitely often, and $\mathbf{x}^*$ is a KKT point;*
2. *There exists an infinite sequence of restoration steps at which the penalty parameter is updated. If $\mathbf{x}^*$ satisfies the LICQ, it is a KKT point; otherwise, it is an FJ point;*
3. *The restoration phase converges to a minimum of the constraint violation.*

BCD Sufficient Decrease

- We establish a generalized Cauchy point with projected gradient descent Beck and Tetruashvili [2013, Lemma 3.3].

$$x_i^{(j+1)} \leftarrow \operatorname*{proj}_{\mathcal{X}_i}\left(x_i^{(j)} - (1/L_i)\nabla_i \mathcal{L}(\boldsymbol{x}_i^{(j)})\right), \quad d_i\left(\boldsymbol{x}_i^{(j)}\right) = x_i^{(j+1)} - x_i^{(j)} \tag{15}$$

Lemma

[Block Coordinate Descent Sufficient Decrease] Let $\{\boldsymbol{x}^{(j)}\}_{j\geq 0}$ be the sequence generated by taking one sequence of projected gradient steps (15) in each block $i = 1, \ldots, p$. Then

$$\mathcal{L}(\boldsymbol{x}^{(j)}) - \mathcal{L}(\boldsymbol{x}^{(j+1)}) \geq \frac{\sigma_{\min}}{2\left[1 + (p-1)\left(\frac{L}{L_{\min}}\right)^2\right]} \left\| d\left(\boldsymbol{x}^{(j)}\right)\right\|^2. \tag{16}$$

for L the Lipschitz constant of $\mathcal{L}(\cdot)$, $L_{\min} = \min_{i\in\{1,\ldots p\}} L_i$, for L_i block Lipschitz constants, and $\sigma_{\min} = \min_{i\in\{1,\ldots p\}} \sigma_i$.

Algorithm Discussion

- ADMM/BCD - solves multiblock problems, parallelizable
  - Requires BCD satisfies a sufficient decrease condition.
- Filter - maintains feasibility, can handle nonlinear constraints, update ρ_k when needed
- Now - applying to bilinear NMF and extensions

Bilinear NMF Problem Formulation

- Extract from a data matrix $M \in \mathbb{R}^{N\times Q}$ two nonnegative factors $X \in \mathbb{R}^{N\times K}$, $Y \in \mathbb{R}^{K\times Q}$

$$\begin{aligned} &\min_{X,Y} \tfrac{1}{2}\|XY - M\|_F^2 \\ &\text{such that } X \geq 0, Y \geq 0 \end{aligned} \tag{17}$$

- ADMM Formulation

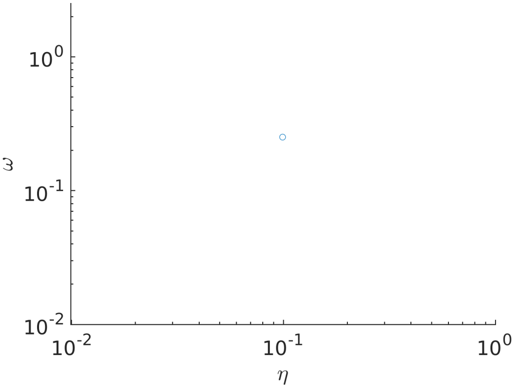
$$\boxed{\begin{aligned} &\min_{X,Y,Z} \tfrac{1}{2}\|Z - M\|_F^2 \\ &\text{such that } X \geq 0, Y \geq 0, Z = XY \end{aligned}} \tag{18}$$

Bilinear NMF Results with ADMM-F

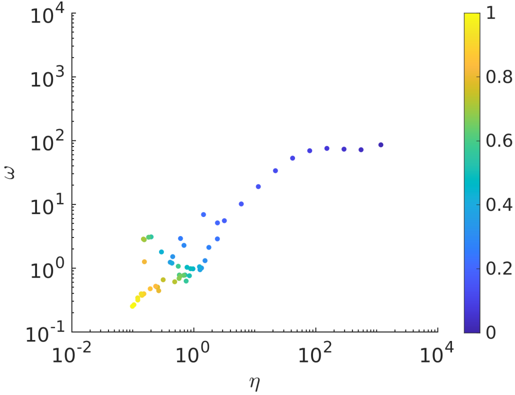
(b) *Data* (c) *XY*

NMF recapitulation of Sylvester and Filter progression through “acceptable” iterations. $M \in \mathbb{R}^{225\times 225}$, $X^{\in}\mathbb{R}^{225\times 45}$. FO:$1e-5$.

Filter Movement



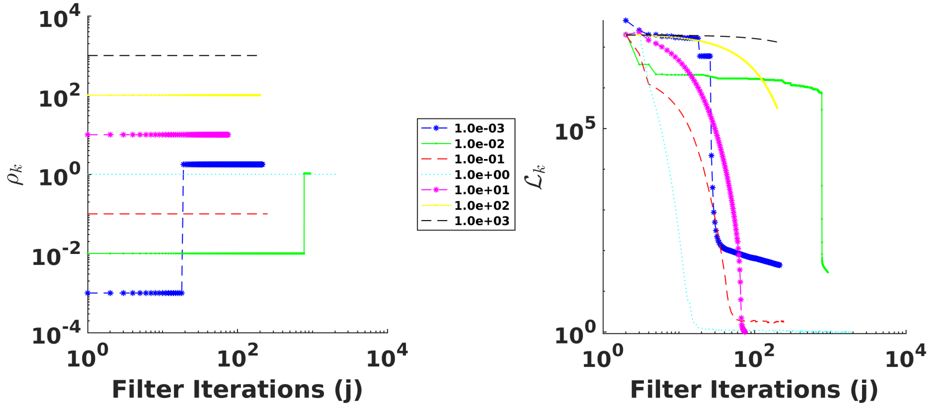
(a) Final Filter



(b) Total Filter Entries with blue earliest, yellow the latest entries.

Descent and Filter Stats NMF for Sylvester T. Cat.

ρ Tests

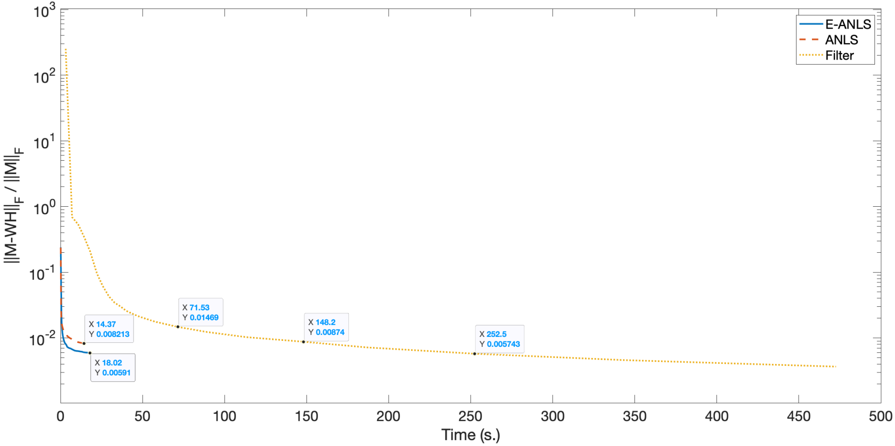


(a) ρ Trajectories

(b) ρ - L_k descent

Convergence metrics for different ρ initializations.

Algorithm Comparison



ADMM-F Comparison with Accelerated NMF Ang and Gillis [2019]

Bilinear NMFC Problem Formulation

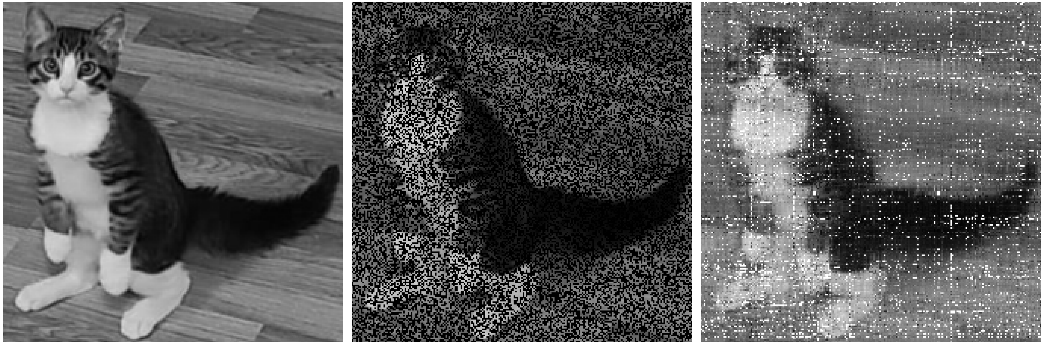
- Data matrix $M \in \mathbb{R}^{N\times Q}$ now contains missing values
- Need two nonnegative factors $X \in \mathbb{R}^{N\times K}$, $Y \in \mathbb{R}^{K\times Q}$ that also match observed data (via selection operator $\mathrm{proj}_{\mathcal{X}}$)

$$\min_{X,Y} \frac{1}{2}\|\operatorname*{proj}_{\mathcal{X}}(XY - M)\|_F^2, \text{such that } X \geq 0, Y \geq 0 \tag{19}$$

- ADMM Formulation

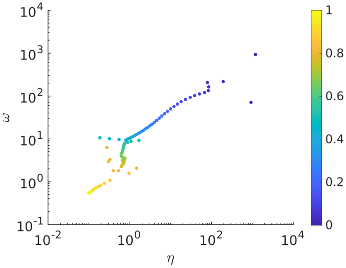
$$\boxed{\begin{aligned} &\min_{X,Y,Z,W} \frac{1}{2}\|Z - W\|_F^2 \\ &\text{such that } X \geq 0, Y \geq 0 \\ &\qquad\qquad Z = XY, \operatorname*{proj}_{\mathcal{X}}(W - M) = 0 \end{aligned}} \tag{20}$$

Bilinear NMFC Results with ADMM-F



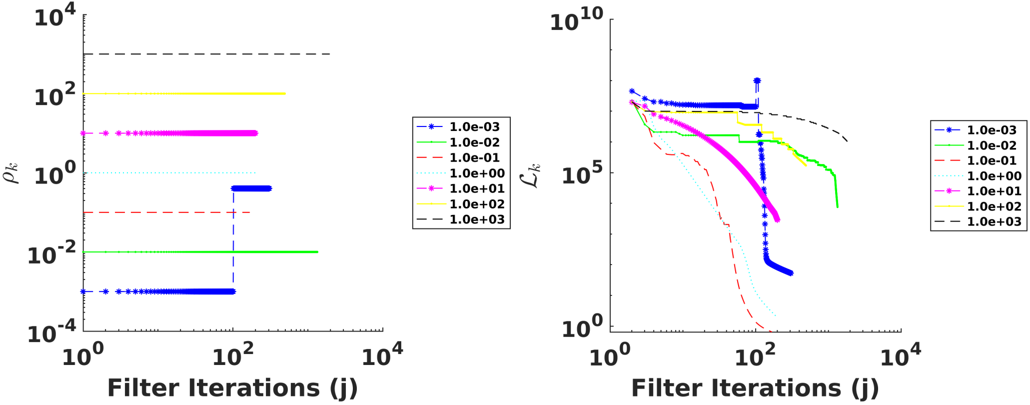
(a) True (b) Data: 50% missing data (c) XY

Bilinear NMFC Results with ADMM-F II



(d) Total Filter Entries: early (blue); final (yellow)

NMFC recapitulation of Sylverster and Filter progression through "acceptable" iterations. $M \in \mathbb{R}^{225\times225}$, $X \in \mathbb{R}^{225\times45}$. FO:$1e-5$.



(a) ρ Trajectories

(b) ρ - L_k descent

Convergence metrics for different ρ initializations.

Chemical Spectrum - Recap

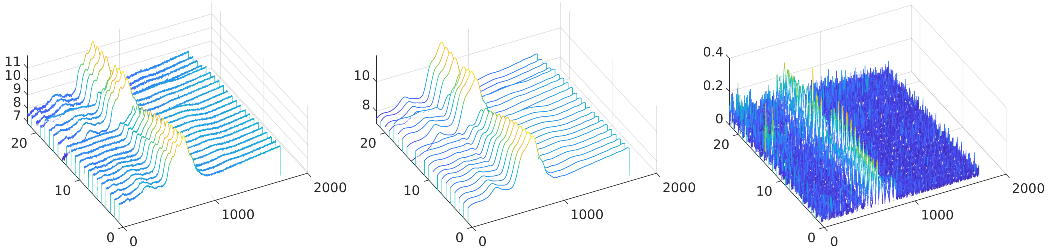
- Find distributions of chemicals that occur in measured spectrum analysis - optimize over μ, σ, Y.
- Each chemical follows a Gaussian distribution; need nonnegative combination of Gaussians that reproduce the spectra data.
- *Intensity Function* as a function of wave number w and concentration c, with rank constraint K. For m be the number of concentrations c:

$$\hat{I}(w_i, c_j) = \sum_{k=1}^{K} \underbrace{h_{k,j}}_{f_k(c_j)} \underbrace{\exp\left(-\frac{1}{2}\left(\frac{w_i - \mu_k}{\sigma_k}\right)^2\right)}_{g_k(w_i)} \tag{21}$$

- Cost Function:

$$\min_{\mu,\sigma,Y} \sum_{i,j} (M_{i,j} - Z)^2 \quad \text{s.t.} \quad Y, \sigma \geq 0, Z = X(\mu, \sigma)Y \tag{22}$$

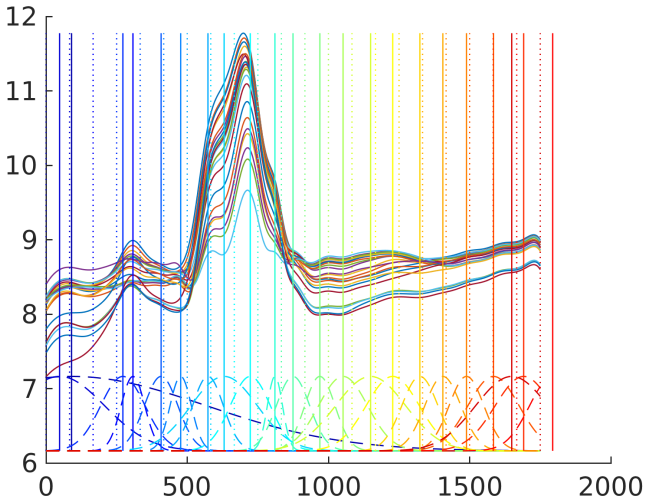
NMF Spectra Results with ADMM-F



(a) Data (b) XY (c) Absolute Difference

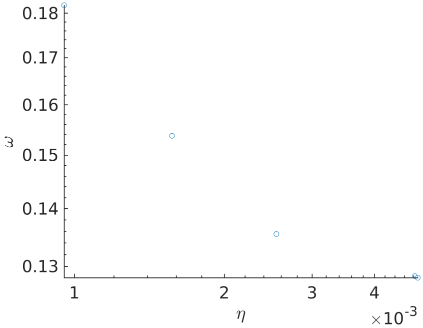
Spectra data and recapitulation through “acceptable” iterations. $M \in \mathbb{R}^{1750\times 22}$, $X \in \mathbb{R}^{22\times 22}$

Distribution of Spectra

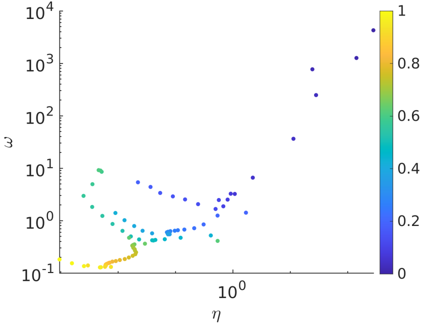


(a) Gaussians, with mean (before ':'/after –) and signal.

Descent Stats

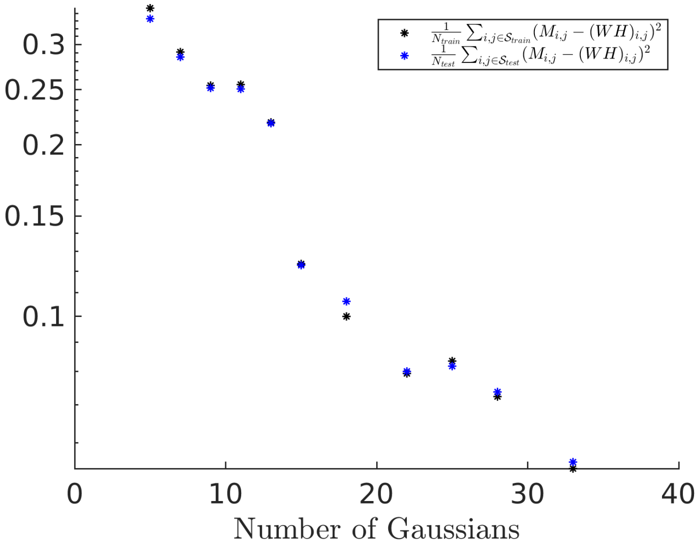


(b) Final Filter



(c) Total Filter Entries

Descent and Filter Stats for spectra inversion with 22 Gaussians.



(22) value per number of Gaussians; training & test data.

Conclusions & Future Directions

- Conclusions
  - Able to prove convergence through Filter framework
  - Filter seems able to select correct ρ values
- Theoretical - Future Directions
  - Prove Convergence w/e nonsmooth regularizers
    - BCD - already have drafted
    - Filter convergence adjustment
    - Accelerations
- Practical
  - Write real package that generalizes to nonlinear problems
  - Incorporate into existing code?

Thank you!

- Questions?

A. M. S. Ang and N. Gillis. Accelerating nonnegative matrix factorization algorithms using extrapolation. *Neural Computation*, 31(2):417–439, 2019. doi: 10.1162/neco_a_01157.

A. Beck and L. Tetruashvili. SIAM journal of optimization. *On the Convergence of Block Coordinate Descent Type Methods*, 23(4), 2013. doi: https://doi.org/10.1137/120887679.

C. Chen, Y. Ye, B.-S. He, and X. Yuan. The direct extension of admm for multi-block convex minimization problems is not necessarily convergent. *Mathematical Programming*, 155:57–79, 2016.

J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American Mathematical Society*, 82:421–439, 1956.

D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2:17–40, 1975.

W. Gao, D. Goldfarb, and F. E. Curtis. Admm for multiaffine constrained optimization. *Optimization Methods and Software*, 35(2):257–303, 2020. doi: 10.1080/10556788.2019.1683553. URL `https://doi.org/10.1080/10556788.2019.1683553`.

R. Glowinski and A. Marroco. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problémes de Dirichlet non linéaires. *ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse Numérique*, 9(R2):41–76, 1975.

C. Vanaret and S. Leyffer. An augmented Lagrangian filter method. *Mathematical Methods of Operations Research*, 2020.

Y. Wang, W. Yin, and J. Zeng. Global convergence of admm in nonconvex nonsmooth optimization. *Journal of Scientific Computing*, 78:29–63, 2018.