Compressed Sensing

Lecture 5
LASSO in Exponential Families
Theory and Algorithms

Alexander Jung

institute of telecommunications

TU WIEN

Recap

- Consider the SLM $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{w}$ with true signal $\mathbf{x} \in \mathcal{X}_s$

- If $\mathbf{w} = \mathbf{0}$ and $\mathrm{spark}(\mathbf{A}) > 2s$ the estimator/recovery scheme

$$\hat{\mathbf{x}}(\mathbf{z}) = \underset{\mathbf{x}'}{\mathrm{argmin}} \, \|\mathbf{x}'\|_0 \text{ subject to } \mathbf{A}\mathbf{x}' = \mathbf{z} \quad \text{(P0)}$$

  is successful

- A computational efficient alternative to P0 is Basis Pursuit (BP)

$$\hat{\mathbf{x}}(\mathbf{z}) = \underset{\mathbf{x}'}{\mathrm{argmin}} \, \|\mathbf{x}'\|_1 \text{ subject to } \|\mathbf{A}\mathbf{x}' - \mathbf{z}\|_2 \leq \eta$$

The LASSO

- By convex duality, BP is equivalent to

$$\hat{\mathbf{x}}^{(\lambda)}(\mathbf{z}) = \operatorname*{argmin}_{\mathbf{x}'} \|\mathbf{z} - \mathbf{A}\mathbf{x}'\|_2^2 + \lambda \|\mathbf{x}'\|_1$$

  for some λ (which may depend on $\mathbf{z}$!)

- In this form, it is called the LASSO estimator

- LASSO stands for "Least Absolute Shrinkage and Selection Operator"

- LASSO corresponds to a convex optimization problem

- Solution path $\hat{\mathbf{x}}^{(\lambda)}(\mathbf{z})$ for whole range of λ can be computed efficiently

Menu

- LASSO is Regularized Maximum Likelihood
- Exponential Families
- Analysis of the LASSO
- Alternating Direction Method of Multipliers (ADMM)

Menu

- LASSO is Regularized Maximum Likelihood
- Exponential Families
- Analysis of the LASSO
- Alternating Direction Method of Multipliers (ADMM)

LASSO for the SLM

- Consider LASSO for the SLM

$$\hat{\mathbf{x}} = \underset{\mathbf{x}'}{\operatorname{argmin}} \|\mathbf{z} - \mathbf{A}\mathbf{x}'\|_2^2 + \lambda\|\mathbf{x}'\|_1$$

- The objective function is composed of two components:
  - the prediction error $\|\mathbf{z} - \mathbf{A}\mathbf{x}'\|_2^2$
  - and the penalty $\lambda\|\mathbf{x}'\|_1$ enforcing a sparse solution $\hat{\mathbf{x}}$

- Likelihood of SLM: $f(\mathbf{z}; \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{z} - \mathbf{A}\mathbf{x}\|_2^2\right)$

- Prediction error is equal (up to constants) to minus log-likelihood

$$\begin{aligned}\log f(\mathbf{z}; \mathbf{x}') &= \log \frac{1}{\sqrt{2\pi\sigma^{2m}}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{z} - \mathbf{A}\mathbf{x}'\|_2^2\right) \\ &= \text{const.} - \frac{1}{2\sigma^2}\|\mathbf{z} - \mathbf{A}\mathbf{x}'\|_2^2\end{aligned}$$

Generalizing LASSO for SLM

- Thus, the LASSO for the SLM can be written as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}'}{\operatorname{argmin}} - \log f(\mathbf{z}; \mathbf{x}') + \lambda \|\mathbf{x}'\|_1$$

- This is ℓ_1 regularized maximum likelihood estimation

- Why not apply LASSO to more general models $f(\mathbf{z}; \mathbf{x})$?

- Allow for general pdfs $f(\mathbf{z}; \mathbf{x})$ $\Rightarrow$ CS with nonlinear measurements

- A wide class of estimation problems can be casted via an exponential family

Menu

- LASSO is Regularized Maximum Likelihood
- Exponential Families
- Analysis of the LASSO
- Alternating Direction Method of Multipliers (ADMM)

Exponential Families

- An exponential family is a parametrized family $\{f(\mathbf{z};\mathbf{x})\}_{\mathbf{x}\in\mathcal{X}}$ of pdfs

$$f(\mathbf{z};\mathbf{x}) = \exp(\mathbf{x}^T\mathbf{\Phi}(\mathbf{z}) - A(\mathbf{x}))h(\mathbf{z}), \text{ for } \mathbf{x}\in\mathcal{X},$$

  with
  - $\mathbf{x}\in\mathcal{X}$: unknown signal (non-linearly) generating observation $\mathbf{z}$
  - $\mathbf{\Phi}(\mathbf{z})$: the vector of sufficient statistics
  - $A(\mathbf{x})$: the cumulant- or log-partition function
  - $h(\mathbf{z})$: a weight function
  - $\mathcal{X}$: the parameter set (corresponding to the signal model, e.g., $\mathcal{X}_s$)

- We require $\mathcal{X}\subseteq\mathcal{N}$, with the natural parameter space

$$\mathcal{N} := \left\{\mathbf{x}\in\mathbb{R}^p : \int_{\mathbf{z}} \exp\left(\mathbf{x}^T\mathbf{\Phi}(\mathbf{z})\right)h(\mathbf{z})d\mathbf{z} < \infty\right\}$$

Examples of Exponential Families: The SLM

- Consider the SLM $\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{w}$ with AWGN $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$

- The pdf $f(\mathbf{z};\mathbf{x}) = \frac{\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{z}-\mathbf{A}\mathbf{x}\|_2^2\right)}{(2\pi\sigma^2)^{m/2}}$ forms exponential family

- The sufficient statistic is given by $\mathbf{\Phi}(\mathbf{z}) = (1/\sigma^2)\mathbf{A}^T\mathbf{z}$

- The cumulant function is $A(\mathbf{x}) = (1/2\sigma^2)\|\mathbf{A}\mathbf{x}\|_2^2$

- The weight function is $h(\mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{m/2}}\exp\left(-(1/2\sigma^2)\|\mathbf{z}\|_2^2\right)$

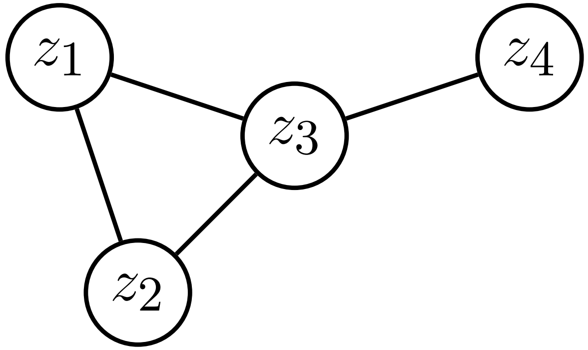
- The parameter set is $\mathcal{X}_s \subseteq \mathcal{N} = \mathbb{R}^p$

Examples of Exponential Families - GMRF

- Consider a Gaussian Markov random field $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with graph $\mathcal{G}$

- The edges in $\mathcal{G}$ correspond to the non-zero entries of $\mathbf{C}^{-1}$

$$\mathbf{C}^{-1} = \begin{pmatrix} 1 & 0.1 & 0.1 & 0 \\ 0.1 & 1 & 0.1 & 0 \\ 0.1 & 0.1 & 1 & 0.1 \\ 0 & 0 & 0.1 & 1 \end{pmatrix}$$



- z_l, z_k connected iff they are conditionally dependent given the rest

- Graphical model selection for GMRF refers to determining edges

Examples of Exponential Families - GMRF (ctd.)

- Let us define the parameter matrix $\mathbf{X} := \mathbf{C}^{-1}$
- Then, $f(\mathbf{z}; \mathbf{X}) = \frac{\sqrt{\det \mathbf{X}}}{(2\pi)^{m/2}} \exp\left(-(1/2)\mathbf{z}^T \mathbf{X} \mathbf{z}\right)$ forms exponential family
- The (matrix-valued) sufficient statistic is $\mathbf{\Phi}(\mathbf{z}) = -(1/2)\mathbf{z}\mathbf{z}^T$
- The cumulant function is $A(\mathbf{X}) = -(1/2) \log \det \mathbf{X}$
- The natural parameter space is $\mathcal{N} = \{\mathbf{X} : \mathbf{X} \text{ is positive definite}\}$
- A sparse graph $\mathcal{G}$ corresponds to sparse matrix $\mathbf{X}$

Properties of Exponential Families

- The sufficient statistic $\mathbf{\Phi}(\mathbf{z})$ carries same information as $\mathbf{z}$

- Thus, we can base our consideration on $\mathbf{t} := \mathbf{\Phi}(\mathbf{z})$ instead of $\mathbf{z}$

- The cumulant function $A(\mathbf{x})$ is a smooth convex function

- The first order partial derivatives satisfy $\frac{\partial A(\mathbf{x})}{\partial x_l} = \mathrm{E}\{t_l\}$

- Hessian of $A(\mathbf{x})$ is given by the Fisher information matrix (FIM) $\mathbf{J}(\mathbf{x})$:

$$\left(\mathbf{J}(\mathbf{x})\right)_{i,j} := \mathrm{E}\left\{ \frac{\partial}{\partial x_i} \log f(\mathbf{z};\mathbf{x}) \frac{\partial}{\partial x_j} \log f(\mathbf{z};\mathbf{x}) \right\} = \mathsf{cov}\{t_i, t_j\}$$

Maximum Likelihood for Exponential Families

- Consider $\mathbf{z}$ following an exp. family $f(\mathbf{z};\mathbf{x})$ with signal model $\mathbf{x} \in \mathcal{X}$

- Maximum Likelihood (ML) estimator of signal $\mathbf{x}$ given by

$$\hat{\mathbf{x}} = \underset{\mathbf{x}' \in \mathcal{X}}{\operatorname{argmax}} \, f(\mathbf{z};\mathbf{x}') = \underset{\mathbf{x}' \in \mathcal{X}}{\operatorname{argmin}} \, -\mathbf{t}^T \mathbf{x}' + A(\mathbf{x}')$$

with sufficient statistic $\mathbf{t} := \mathbf{\Phi}(\mathbf{z})$

- Thus, ML corresponds to a convex optimization problem

- Typically, $\mathbf{t} \approx \boldsymbol{\mu} := \mathrm{E}\{\mathbf{\Phi}(\mathbf{z})\}$ where expect. is w.r.t. to $f(\mathbf{z};\mathbf{x})$

- It can be shown that $\nabla(-\boldsymbol{\mu}^T \mathbf{x}' + A(\mathbf{x}'))\big|_{\mathbf{x}'=\mathbf{x}} = \mathbf{0}$

Analysis of ML for Exponential Families

- ML estimator of signal $\mathbf{x}$ given by

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}' \in \mathcal{X}} -(\mathbf{t} - \boldsymbol{\mu})^T \mathbf{x}' - \boldsymbol{\mu}^T \mathbf{x}' + A(\mathbf{x}')$$

- Since $\nabla(-\boldsymbol{\mu}^T \mathbf{x}' + A(\mathbf{x}'))\big|_{\mathbf{x}'=\mathbf{x}} = \mathbf{0}$, the quadratic approx. yields

$$-\boldsymbol{\mu}^T \mathbf{x}' + A(\mathbf{x}') \approx \text{const.} + (1/2)(\mathbf{x}' - \mathbf{x})^T \mathbf{J}(\mathbf{x}' - \mathbf{x})$$

- Setting gradient in ML objective to zero yields $\hat{\mathbf{x}} - \mathbf{x} \approx \mathbf{J}^{-1}(\mathbf{x})(\mathbf{t} - \boldsymbol{\mu})$

- Estimation error $\hat{\mathbf{x}} - \mathbf{x}$ is governed by two parts:
  - deterministic part due to condition of FIM $\mathbf{J}(\mathbf{x})$
  - stochastic part due to deviation of $\mathbf{t}(\mathbf{z})$ from its mean $\boldsymbol{\mu} = \mathrm{E}\{\mathbf{t}\}$

Analysis of ML - Deterministic vs. Stochastic Part

- The deterministic part of the ML analysis is characterizing $\mathbf{J}(\mathbf{x})$ for $\mathbf{x} \in \mathcal{X}$
- Here, the signal model $\mathcal{X}$ comes into play!
- Stochastic part: characterize the deviation of $\mathbf{t} = \mathbf{\Phi}(\mathbf{z})$ from its mean
- The stochastic part does not take the signal model into account!
- We need large deviation analysis in the form of tails bounds

$$\mathrm{P}\{\|\mathbf{t} - \boldsymbol{\mu}\|_\infty \geq \varepsilon\} \leq ?,$$

with $\boldsymbol{\mu} = \mathrm{E}\{\mathbf{t}\}$

Large Deviations of Exponential Families

- Given an exponential family $f(\mathbf{z};\mathbf{x}) = \exp(\mathbf{x}^T\mathbf{\Phi}(\mathbf{z}) - A(\mathbf{x}))h(\mathbf{z})$

- Consider sufficient statistic $\mathbf{t} = \mathbf{\Phi}(\mathbf{z})$ with mean $\boldsymbol{\mu} = \mathrm{E}\{\mathbf{\Phi}(\mathbf{z})\}$

- It can be shown that

$$\mathrm{P}\{|t_l - \mu_l| \geq \varepsilon\} \leq c\exp(-r\varepsilon)$$

  where c and r depend only on the cumulant function $A(\mathbf{x})$

LASSO for Exponential Families

- Consider an exponential family $f(\mathbf{z};\mathbf{x}) = \exp(\mathbf{x}^T\mathbf{\Phi}(\mathbf{z}) - A(\mathbf{x}))h(\mathbf{z})$

- The unknown signal $\mathbf{x}$ is known to be sparse, i.e., $\|\mathbf{x}\|_0 \leq s$ or $\mathcal{X} = \mathcal{X}_s$

- Enforce sparsity by adding $\lambda\|\mathbf{x}'\|_1$ to ML objective instead of using constraint $\mathbf{x}' \in \mathcal{X}_s$

- This yields the LASSO for a general exponential family

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}' \in \mathcal{N}} -\mathbf{t}^T\mathbf{x}' + A(\mathbf{x}') + \lambda\|\mathbf{x}'\|_1$$

Menu

- LASSO is Regularized Maximum Likelihood
- Exponential Families
- Analysis of the LASSO
- Alternating Direction Method of Multipliers (ADMM)

Characterizing Solution of a Minimization

- Consider a general unconstrained minimization $\hat{\mathbf{x}} = \mathrm{argmin}_{\mathbf{x}' \in \mathbb{R}^p} f(\mathbf{x}')$
- How can we characterize $\hat{\mathbf{x}}$?
- For differentiable function $f(\cdot)$, necessary condition: $\nabla f(\hat{\mathbf{x}}) = \mathbf{0}$
- For convex function $f(\cdot)$, which may be non-differentiable,

$$\mathbf{0} \in \partial f(\hat{\mathbf{x}})$$

  where $\partial f(\hat{\mathbf{x}})$ denotes subdifferential of $f(\cdot)$ at $\hat{\mathbf{x}}$

- For an arbitrary function $f(\cdot)$:

$$f(\hat{\mathbf{x}}) \leq f(\mathbf{x}'), \text{ for all } \mathbf{x}'$$

LASSO yields Sparse Results - I

- Consider LASSO $\hat{\mathbf{x}}$ and true signal $\mathbf{x}$ generating measurements via $f(\mathbf{z};\mathbf{x})$

- By definition of LASSO we have the Basic Inequality

$$-\mathbf{t}^T\hat{\mathbf{x}} + A(\hat{\mathbf{x}}) + \lambda\|\hat{\mathbf{x}}\|_1 \leq -\mathbf{t}^T\mathbf{x} + A(\mathbf{x}) + \lambda\|\mathbf{x}\|_1$$

- Let's rewrite this using deviation of suff. statistic $\mathbf{t}$ from its mean $\boldsymbol{\mu}$:

$$-(\mathbf{t}-\boldsymbol{\mu})^T\hat{\mathbf{x}}-\boldsymbol{\mu}^T\hat{\mathbf{x}}+A(\hat{\mathbf{x}})+\lambda\|\hat{\mathbf{x}}\|_1 \leq -(\mathbf{t}-\boldsymbol{\mu})^T\mathbf{x}-\boldsymbol{\mu}^T\mathbf{x}+A(\mathbf{x})+\lambda\|\mathbf{x}\|_1$$

- Since $\operatorname{argmin}_{\mathbf{x}'} -\boldsymbol{\mu}^T\mathbf{x}' + A(\mathbf{x}') = \mathbf{x}$, we obtain further:

$$-(\mathbf{t}-\boldsymbol{\mu})^T\hat{\mathbf{x}} + \lambda\|\hat{\mathbf{x}}\|_1 \leq -(\mathbf{t}-\boldsymbol{\mu})^T\mathbf{x} + \lambda\|\mathbf{x}\|_1$$

LASSO yields Sparse Results - II

- Defer stochastic part by assuming for the moment that $\|\mathbf{t}-\boldsymbol{\mu}\|_\infty \leq \lambda/2$

- This yields
$$\|\hat{\mathbf{x}}\|_1 \leq (1/2)\|\hat{\mathbf{x}}-\mathbf{x}\|_1 + \|\mathbf{x}\|_1$$

- Denoting $\mathcal{S} := \mathrm{supp}(\mathbf{x})$, we lower bound the LHS as
$$\|\hat{\mathbf{x}}\|_1 = \|\hat{\mathbf{x}}_{\mathcal{S}}\|_1 + \|\hat{\mathbf{x}}_{\mathcal{S}^c}\|_1 \geq \|\mathbf{x}_{\mathcal{S}}\| - \|\mathbf{x}_{\mathcal{S}}-\hat{\mathbf{x}}_{\mathcal{S}}\|_1 + \|\hat{\mathbf{x}}_{\mathcal{S}^c}\|_1$$
and upper bound the RHS as
$$\|\hat{\mathbf{x}}-\mathbf{x}\|_1 \leq \|\hat{\mathbf{x}}_{\mathcal{S}^c}\|_1 + \|\mathbf{x}_{\mathcal{S}}-\hat{\mathbf{x}}_{\mathcal{S}}\|_1$$

- Using $\mathbf{x}=\mathbf{x}_{\mathcal{S}}$, we arrive at
$$\|\hat{\mathbf{x}}_{\mathcal{S}^c}\|_1 \leq 3\|\mathbf{x}_{\mathcal{S}}-\hat{\mathbf{x}}_{\mathcal{S}}\|_1$$

Analysis of LASSO

- Consider following form of Basic Inequality

$$-(\mathbf{t}-\boldsymbol{\mu})^T\hat{\mathbf{x}}-\boldsymbol{\mu}^T\hat{\mathbf{x}}+A(\hat{\mathbf{x}})+\lambda\|\hat{\mathbf{x}}\|_1 \le -(\mathbf{t}-\boldsymbol{\mu})^T\mathbf{x}-\boldsymbol{\mu}^T\mathbf{x}+A(\mathbf{x})+\lambda\|\mathbf{x}\|_1$$

- Using $\|\mathbf{t}-\boldsymbol{\mu}\|_\infty \le \lambda/2$, this yields

$$f(\hat{\mathbf{x}}) \le (3/2)\lambda\|\hat{\mathbf{x}}-\mathbf{x}\|_1$$

where $f(\hat{\mathbf{x}}) := -\boldsymbol{\mu}^T(\hat{\mathbf{x}}-\mathbf{x}) + A(\hat{\mathbf{x}}) - A(\mathbf{x})$

- We have $f(\mathbf{x}) = 0$, $\nabla f(\mathbf{x}) = \mathbf{0}$ and Hessian $\nabla^2 f(\mathbf{x}) = \mathbf{J}(\mathbf{x})$

- By Taylor's theorem

$$\begin{aligned} f(\hat{\mathbf{x}}) &= f(\mathbf{x}) + \left(\nabla f(\mathbf{x})\right)^T(\hat{\mathbf{x}}-\mathbf{x}) + (1/2)(\hat{\mathbf{x}}-\mathbf{x})^T\mathbf{J}(\mathbf{x}')(\hat{\mathbf{x}}-\mathbf{x}) \\ &= (1/2)(\hat{\mathbf{x}}-\mathbf{x})^T\mathbf{J}(\mathbf{x}')(\hat{\mathbf{x}}-\mathbf{x}) \end{aligned}$$

where $\mathbf{x}'$ is from the line segment connecting $\hat{\mathbf{x}}$ and $\mathbf{x}$

Compatibility Condition

- Given $\|\mathbf{t}-\boldsymbol{\mu}\|_\infty \leq \lambda/2$, we obtained from Basic Inequality that

$$(\hat{\mathbf{x}}-\mathbf{x})^T\mathbf{J}(\hat{\mathbf{x}}-\mathbf{x}) \leq 3\lambda\|\hat{\mathbf{x}}-\mathbf{x}\|_1$$

- Compatibility condition for $\mathcal{S} := \mathrm{supp}(\mathbf{x})$ with constant ϕ:

  $$(\mathbf{x}')^T\mathbf{J}\mathbf{x}' \geq \phi\|\mathbf{x}'\|_1^2/s \quad \text{for any } \mathbf{x}', \text{ with } \|\mathbf{x}'_{\mathcal{S}^c}\| \leq 3\|\mathbf{x}'_{\mathcal{S}}\|$$

- Note that LASSO error $\mathbf{e} = \hat{\mathbf{x}}-\mathbf{x}$ satisfies $\|\mathbf{e}_{\mathcal{S}^c}\| \leq 3\|\mathbf{e}_{\mathcal{S}}\|$

- If compatibility condition holds, we arrive at

$$\|\hat{\mathbf{x}}-\mathbf{x}\|_1 \leq 3\lambda s/\phi$$

How Strict is the Compatibility Condition ?

- The compatibility condition is very similar to RIP
- Requires small $s \times s$ sub-matrix $\mathbf{J}_{\mathcal{S},\mathcal{S}}$ to be well conditioned
- Without sparsity constraints, whole $p \times p$ matrix $\mathbf{J}$ required well conditioned
- Assume FIM $\mathbf{J}$ is generated as $\mathbf{A}^T\mathbf{A}$ with i.i.d. Gaussian matrix $\mathbf{A}$
- Prob. that $\lambda_{\text{min}}(\mathbf{J}_{\mathcal{S},\mathcal{S}}) \le \delta$ upper bounded by $c_1 \exp(-c_2(1-\delta)^2 p + c_3 s)$
- Prob. that $\lambda_{\text{min}}(\mathbf{J}) \le \delta$ upper bounded by $c_1 \exp(-c_2(1-\delta)^2 p + c_3 p)$

The Final Result

- Consider LASSO for a given exponential family:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}' \in \mathbb{R}^p}{\operatorname{argmin}} -\mathbf{t}^T \mathbf{x}' + A(\mathbf{x}') + \lambda \|\mathbf{x}'\|_1$$

- If $\|\mathbf{t} - \boldsymbol{\mu}\|_\infty \leq \lambda/2$ and compatibility condition holds, we derived that

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq 3\lambda s/\phi$$

- For smaller λ the estimation error tends to be smaller

- However, for too small λ, the condition $\|\mathbf{t} - \boldsymbol{\mu}\|_\infty \leq \lambda/2$ may be violated

- Large values of the compatibility constant ϕ are desirable

Menu

- LASSO is Regularized Maximum Likelihood
- Exponential Families
- Analysis of the LASSO
- Alternating Direction Method of Multipliers (ADMM)

Methods for Computing LASSO

- Consider LASSO for a given exponential family:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}'}{\operatorname{argmin}} -\mathbf{t}^T\mathbf{x}' + A(\mathbf{x}') + \lambda\|\mathbf{x}'\|_1$$

- This is an unconstrained convex optimization problem since $A(\cdot)$ is convex

- Introducing auxiliary variable $\mathbf{u}$, we can rewrite LASSO as

$$\hat{\mathbf{x}} = \underset{\mathbf{x}',\mathbf{u}}{\operatorname{argmin}} -\mathbf{t}^T\mathbf{x}' + A(\mathbf{x}') + \lambda\|\mathbf{u}\|_1$$

$$\text{subject to} \quad \mathbf{u} = \mathbf{x}'$$

- Note splitting of the diff. and non-diff. components in the objective function

Alternating Direction Method of Multipliers (ADMM)

- Another equivalent form of LASSO is

$$\hat{\mathbf{x}} = \underset{\mathbf{x}',\mathbf{u}}{\operatorname{argmin}} -\mathbf{t}^T\mathbf{x}' + A(\mathbf{x}') + \lambda\|\mathbf{u}\|_1 + (\rho/2)\|\mathbf{u}-\mathbf{x}'\|_2^2$$

$$\text{subject to} \quad \mathbf{u} = \mathbf{x}'$$

  for some $\rho > 0$

- The Lagrangian for this optimization problem is

$$L_\rho(\mathbf{x}',\mathbf{u},\mathbf{y}) := -\mathbf{t}^T\mathbf{x}' + A(\mathbf{x}') + \lambda\|\mathbf{u}\|_1 + \mathbf{y}^T(\mathbf{u}-\mathbf{x}') + (\rho/2)\|\mathbf{u}-\mathbf{x}'\|_2^2$$

- ADMM iteratively optimizes Lagrangian $L_\rho(\mathbf{x}',\mathbf{u},\mathbf{y})$

- ADMM breaks original optimization into smaller (easier) subproblems

ADMM for LASSO

- Consider Lagrangian for ℓ_1 regularized ML (LASSO) for an exp. family

  $L_\rho(\mathbf{x}', \mathbf{u}, \mathbf{y}) := -\mathbf{t}^T\mathbf{x}' + A(\mathbf{x}') + \lambda\|\mathbf{u}\|_1 + \mathbf{y}^T(\mathbf{u} - \mathbf{x}') + (\rho/2)\|\mathbf{u} - \mathbf{x}'\|_2^2$

- ADMM constructs sequence $(\mathbf{x}^{(k)}, \mathbf{u}^{(k)}, \mathbf{y}^{(k)})$ by iterating

$$\mathbf{x}^{(k+1)} := \operatorname*{argmin}_{\mathbf{x}'} L_\rho(\mathbf{x}', \mathbf{u}^{(k)}, \mathbf{y}^{(k)})$$

$$\mathbf{u}^{(k+1)} := \operatorname*{argmin}_{\mathbf{u}} L_\rho(\mathbf{x}^{(k+1)}, \mathbf{u}, \mathbf{y}^{(k)})$$

$$\mathbf{y}^{(k+1)} := \mathbf{y}^{(k)} + \rho(\mathbf{x}^{(k+1)} - \mathbf{u}^{(k+1)})$$

- Under mild conditions, iterates converge to LASSO, i.e., $\lim_{k\to\infty} \mathbf{x}^{(k)} = \hat{\mathbf{x}}$

ADMM for LASSO in SLM

- Consider LASSO for the SLM

- Corresponds to exp. family with $A(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|_2^2$ and $\mathbf{t} = (1/2\sigma^2)\mathbf{A}^T\mathbf{z}$

- ADMM iterates are given by

$$\mathbf{x}^{(k+1)} = (\mathbf{A}^T\mathbf{A} + \rho\mathbf{I})^{-1}(\mathbf{A}^T\mathbf{z} + \rho(\mathbf{u}^{(k)} - \mathbf{y}^{(k)}))$$

$$\mathbf{u}^{(k+1)} = \mathcal{S}_{\lambda/\rho}(\mathbf{x}^{(k+1)} + \mathbf{y}^{(k)})$$

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

with the element-wise soft-thresholding operator $\mathcal{S}_\kappa(a) := (1 - \kappa/|a|)_+ a$

ADMM for LASSO in GMRF

- Consider LASSO for a GMRF $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ with "signal matrix" $\mathbf{X} := \mathbf{C}^{-1}$

- Corresponds to exp. family with $A(\mathbf{X}) = -(1/2) \log\det \mathbf{X}$

- ADMM iterates are given by

$$\mathbf{X}^{(k+1)} = \operatorname*{argmin}_{\mathbf{X} \succ \mathbf{0}} \mathbf{z}^T \mathbf{X} \mathbf{z} - \log\det \mathbf{X} + (\rho/2) \|\mathbf{X} - \mathbf{U}^{(k)} + (1/\rho) \mathbf{Y}^{(k)}\|_{\mathsf{F}}^2$$

$$\mathbf{U}^{(k+1)} = \mathcal{S}_{\lambda/\rho}(\mathbf{X}^{(k+1)} + (1/\rho) \mathbf{Y}^{(k)})$$

$$\mathbf{Y}^{(k+1)} = \mathbf{Y}^{(k)} + \rho(\mathbf{X}^{(k+1)} - \mathbf{U}^{(k+1)})$$

- Closed form solution for $\mathbf{X}$ minimization via eigenvalue decomposition

What we learned today

- LASSO for Exponential Families is ℓ_1-regularized ML estimation
- Simple analysis by separation into deterministic and stochastic part
- Stochastic part is a large deviation analysis of sufficient statistic $\mathbf{t} = \mathbf{\Phi}(\mathbf{z})$
- Deterministic part based on "compatibility condition" of FIM
- Efficient implementation of LASSO via ADMM

Happy Easter Holidays!

- See you again after easter holidays on Monday, 28th of April!