UNIVERSITY OF MINNESOTA
Driven to Discover℠

EE 5561: Image Processing and Applications

Lecture 12

Mehmet Akçakaya

Recap of Last Lecture

- Sparsity regularization
  - l_1 norm, soft thresholding
  - Solving the objective function

$$\arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \tau||\mathbf{x}||_1$$

degradation/measurement matrix

  - Gradient descent
  - Proximal gradient descent
- Today:
  - Variable splitting

Lagrangian relaxation

- Some more optimization
  - Suppose we are solving a constrained optimization problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \qquad \text{s. t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

  - The Lagrangian is given by

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b})$$

  - If $\hat{\mathrm{x}}$ is a solution to the original problem, then there exists $\hat{\boldsymbol{\lambda}}$ such that partial derivatives of $\mathcal{L}$ at $(\hat{\mathrm{x}}, \hat{\boldsymbol{\lambda}})$ are 0.

  - For convex *f*, there exists $\hat{\boldsymbol{\lambda}}$ such that $\hat{\mathrm{x}}$ minimizes $\mathcal{L}(\mathbf{x}, \hat{\boldsymbol{\lambda}})$ and $\hat{\boldsymbol{\lambda}}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = 0$

Lagrangian relaxation

- Some more optimization
  - We can define a dual function $g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda})$
  - The dual problem is to solve

$$\hat{\boldsymbol{\lambda}} = \max_{\boldsymbol{\lambda}} g(\boldsymbol{\lambda})$$

  - While we also solve $\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \hat{\boldsymbol{\lambda}})$
  - One can solve this by solving the dual problem by gradient method (ascent)

$$\boldsymbol{\lambda}^{(t+1)} = \boldsymbol{\lambda}^{(t)} + \alpha_k(\mathbf{A}\mathbf{x}^{(t)} - \mathbf{b})$$
$$\mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^{(t+1)})$$

  - Alternating method to solve it (need assumption to work)

Augmented Lagrangian

- Some more optimization
  - Again start from the constrained problem

$$\min_{\mathbf{x}} f(\mathbf{x}) \qquad \text{s. t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

  - Augmented Lagrangian is given by

$$\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$$

augmentation
ensures dual function
$g(\boldsymbol{\lambda}) = \inf_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda})$ is differentiable

Augmented Lagrangian

- Some more optimization
  - Alternative intuition

$$\min_{\mathbf{x}} f(\mathbf{x}) \qquad \text{s. t. } \mathbf{A}\mathbf{x} = \mathbf{b}$$

replace with quadratic penalty
i.e. $||\mathbf{b} - \mathbf{A}\mathbf{x}||_2^2$ should be small

$$f(\mathbf{x}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$$

  - We can do the same dual solution idea for the augmented Lagrangian

$$\mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}^{(t)})$$

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

$$\mathcal{L}_\rho(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} - \mathbf{b}||_2^2$$

  - This is called method of multipliers
    - The first problem is usually hard to solve

ADMM

- Alternating direction method of multipliers (ADMM)
  - Now consider something more similar to last lecture

$$\min_{\mathbf{x},\mathbf{z}} f(\mathbf{x}) + h(\mathbf{z}) \qquad \text{s. t. } \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} = \mathbf{c}$$

    - Difference to last lecture → two variables **x** and **z**
  - Augmented Lagrangian is

$$\mathcal{L}_\rho(\mathbf{x}, \mathbf{z}, \boldsymbol{\lambda}) = f(\mathbf{x}) + h(\mathbf{z}) + \boldsymbol{\lambda}^T(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}) + \frac{\rho}{2}||\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z} - \mathbf{c}||_2^2$$

  - ADMM solves

$$\begin{aligned}
\mathbf{x}^{(t+1)} &= \arg\min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^{(t)}, \boldsymbol{\lambda}^{(t)}) \\
\mathbf{z}^{(t+1)} &= \arg\min_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{z}, \boldsymbol{\lambda}^{(t)}) \\
\boldsymbol{\lambda}^{(t+1)} &= \boldsymbol{\lambda}^{(t)} + \rho(\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{z}^{(t+1)} - \mathbf{c})
\end{aligned}$$

ADMM

- Alternating direction method of multipliers (ADMM)
  - We can make this nicer, consider

$$\boldsymbol{\lambda}^T(\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}) + \frac{\rho}{2}||\mathbf{Ax} + \mathbf{Bz} - \mathbf{c}||_2^2$$

  - Define $\mathbf{r} = \mathbf{Ax} + \mathbf{Bz} - \mathbf{c}$

$$\boldsymbol{\lambda}^T\mathbf{r} + \frac{\rho}{2}||\mathbf{r}||_2^2 = \frac{\rho}{2}\left(\mathbf{r}^T\mathbf{r} + \frac{2}{\rho}\boldsymbol{\lambda}^T\mathbf{r}\right)$$

$$= \frac{\rho}{2}\left(\mathbf{r}^T\mathbf{r} + 2\frac{\boldsymbol{\lambda}^T}{\rho}\mathbf{r} + \frac{\boldsymbol{\lambda}^T\boldsymbol{\lambda}}{\rho^2} - \frac{\boldsymbol{\lambda}^T\boldsymbol{\lambda}}{\rho^2}\right) = \frac{\rho}{2}\left(\left|\left|\mathbf{r} + \frac{\boldsymbol{\lambda}}{\rho}\right|\right|_2^2 - \left|\left|\frac{\boldsymbol{\lambda}}{\rho}\right|\right|_2^2\right)$$

  - Define $\mathbf{u} = \boldsymbol{\lambda}/\rho$

$$= \frac{\rho}{2}\left(||\mathbf{r} + \mathbf{u}||_2^2 - ||\mathbf{u}||_2^2\right)$$

ADMM

- Alternating direction method of multipliers (ADMM)
  - Now the updates become

$$
\begin{aligned}
\mathbf{x}^{(t+1)} &= \arg\min_{\mathbf{x}} \mathcal{L}_\rho(\mathbf{x}, \mathbf{z}^{(t)}, \mathbf{u}^{(t)}) \\
&= \arg\min_{\mathbf{x}} f(\mathbf{x}) + \frac{\rho}{2} ||\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}^{(t)} - \mathbf{c} + \mathbf{u}^{(t)}||_2^2 \\
\mathbf{z}^{(t+1)} &= \arg\min_{\mathbf{z}} \mathcal{L}_\rho(\mathbf{x}^{(t+1)}, \mathbf{z}, \mathbf{u}^{(t)}) \\
&= \arg\min_{\mathbf{z}} h(\mathbf{z}) + \frac{\rho}{2} ||\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{z} - \mathbf{c} + \mathbf{u}^{(t)}||_2^2 \\
\mathbf{u}^{(t+1)} &= \mathbf{u}^{(t)} + (\mathbf{A}\mathbf{x}^{(t+1)} + \mathbf{B}\mathbf{z}^{(t+1)} - \mathbf{c})
\end{aligned}
$$

ADMM in Computational Imaging

- Back to computational imaging

$$\arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{x}) \qquad \text{e.g. } \psi(\mathbf{x}) = \tau||\mathbf{W}\mathbf{x}||_1$$

  - This is unconstrained, so not the same form as ADMM
  - But we can do “variable splitting”

$$\arg\min_{\mathbf{x},\mathbf{z}} \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{z}) \quad \text{s. t.} \quad \mathbf{x} = \mathbf{z}$$

  - Then we have

$$\mathcal{L}_\rho(\mathbf{x},\mathbf{z},\mathbf{u}) = \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{z}) + \frac{\rho}{2}||\mathbf{x} - \mathbf{z} + \mathbf{u}||_2^2$$

$$\mathcal{L}_\rho(\mathbf{x},\mathbf{z},\mathbf{u}) = \underbrace{\frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2}_{f(\mathbf{x})} + \underbrace{\psi(\mathbf{z})}_{h(\mathbf{z})} + \underbrace{\frac{\rho}{2}||\mathbf{x} - \mathbf{z} + \mathbf{u}||_2^2}_{\text{constraint + Lagrangian}}$$

ADMM in Computational Imaging

- Compare this to the previous "intuition" where we replace the constraint **x** = **z** with $||\mathbf{x}-\mathbf{z}||_2^2$ should be small

$$\frac{1}{2}||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{z}) + \frac{\rho}{2}||\mathbf{x}-\mathbf{z}||_2^2$$

variable splitting with quadratic penalty

- Looks similar to $\mathcal{L}_\rho(\mathbf{x},\mathbf{z},\mathbf{u}) = \frac{1}{2}||\mathbf{y}-\mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{z}) + \frac{\rho}{2}||\mathbf{x}-\mathbf{z}+\mathbf{u}||_2^2$
- But the **u** term makes ADMM converge faster

- One more point about ADMM
  - Has slow convergence to high accuracy
  - Fast convergence to medium accuracy
  - Latter is sufficient for image processing

ADMM in Computational Imaging

- Going back to ADMM, the updates are

i) $$\mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \frac{\rho}{2}||\mathbf{x} - \mathbf{z}^{(t)} + \mathbf{u}^{(t)}||_2^2$$
$$= \left(\mathbf{H}^T\mathbf{H} + \rho\mathbf{I}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \rho(\mathbf{z}^{(t)} - \mathbf{u}^{(t)})\right)$$

ii) $$\mathbf{z}^{(t+1)} = \arg\min_{\mathbf{z}} h(\mathbf{z}) + \frac{\rho}{2}||\mathbf{x}^{(t+1)} - \mathbf{z} + \mathbf{u}^{(t)}||_2^2$$
$$= \arg\min_{\mathbf{z}} \frac{1}{2}||(\mathbf{x}^{(t+1)} + \mathbf{u}^{(t)}) - \mathbf{z}||_2^2 + \frac{1}{\rho}\psi(\mathbf{z})$$

"denoising" or proximal operator

For $\psi(\mathbf{z}) = \tau||\mathbf{z}||_1$, solution is $$\mathbf{z}^{(t+1)} = \arg\min_{\mathbf{z}} \frac{1}{2}||(\mathbf{x}^{(t+1)} + \mathbf{u}^{(t)}) - \mathbf{z}||_2^2 + \frac{\tau}{\rho}||\mathbf{z}||_1$$
$$= S_{\frac{\tau}{\rho}}(\mathbf{x}^{(t+1)} + \mathbf{u}^{(t)})$$

iii) $$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + (\mathbf{x}^{(t+1)} - \mathbf{z}^{(t+1)})$$

Solving Regularized Least Squares

- Overall we have two algorithms for solving $\arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi(\mathbf{x})$
  - Proximal gradient descent

$$\mathbf{z}^{(t+1)} = \mathbf{x}^{(t)} + \eta\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(t)})$$

$$\mathbf{x}^{(t+1)} = \arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{z}^{(t+1)} - \mathbf{x}||_2^2 + \eta\psi(\mathbf{x})$$

"denoising" or proximal operator

  - ADMM

$$\mathbf{x}^{(t+1)} = \left(\mathbf{H}^T\mathbf{H} + \rho\mathbf{I}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \rho(\mathbf{z}^{(t)} - \mathbf{u}^{(t)})\right)$$

$$\mathbf{z}^{(t+1)} = \arg\min_{\mathbf{z}} \frac{1}{2}||(\mathbf{x}^{(t+1)} + \mathbf{u}^{(t)}) - \mathbf{z}||_2^2 + \frac{1}{\rho}\psi(\mathbf{z})$$

"denoising" or proximal operator

$$\mathbf{u}^{(t+1)} = \mathbf{u}^{(t)} + (\mathbf{x}^{(t+1)} - \mathbf{z}^{(t+1)})$$

Solving Regularized Least Squares

- Both converge when $\psi(\mathbf{x}) = \tau||\mathbf{x}||_1$
  - Proximal operator easy to solve in this case → soft thresholding

- In reality, one can use other $\psi(\cdot)$ (or equivalently $\mathrm{prox}_{\psi}(\cdot)$), sometimes without guarantees
  - e.g. replace proximal operator patch-based denoising methods (next lecture)

Solving Regularized Least Squares

- ADMM is more versatile
  - Can add multiple regularizers, e.g.

$$\arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi_1(\mathbf{x}) + \psi_2(\mathbf{x})$$

  - Then variable splitting is done as

$$\arg\min_{\mathbf{x},\mathbf{z},\mathbf{v}} \frac{1}{2}||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \psi_1(\mathbf{z}) + \psi_2(\mathbf{v}) \qquad \text{s. t. } \mathbf{x} = \mathbf{z} \text{ and } \mathbf{x} = \mathbf{v}$$

  - Not as easy in proximal gradient descent

Solving Regularized Least Squares

- Final point: How do we solve

$$\mathbf{x}^{(t+1)} = \left(\mathbf{H}^T\mathbf{H} + \rho\mathbf{I}\right)^{-1}\left(\mathbf{H}^T\mathbf{y} + \rho(\mathbf{z}^{(t)} - \mathbf{u}^{(t)})\right)$$

  - In practice, we cannot invert (or even store) these matrices
  - These least squares problems are themselves solved iteratively
  - Again, gradient descent is simplest, but slow
  - Practically, one uses the conjugate gradient method (works if **A** is symmetric and positive definite)
    - It is in the ADMM subproblem $\mathbf{A} = (\mathbf{H}^T\mathbf{H} + \rho\mathbf{I})$
    - Also in (generalized) Tikhonov regularization
    - We are not going to cover why this works
    - But this is the function we will call in our implementations

Example Solution

Original

Blurred w/ Noise

Recovered