Proximal mapping via network optimization

- minimum cut and maximum flow problems
- parametric minimum cut problem
- application to proximal mapping

Introduction

this lecture: network flow algorithm for evaluating prox-operator of

$$
\begin{aligned}
h(x) &= \sum_{i=1}^{n} \sum_{j=1}^{i-1} A_{ij} |x_i - x_j| \\
&= \sum_{i,j=1}^{n} A_{ij} \max\{0, x_i - x_j\}
\end{aligned}
$$

- coefficients $A_{ij} = A_{ji}$ are nonnegative
- associated undirected graph has n nodes, edges (i, j) when $A_{ij} > 0$
- applications in image processing and machine learning

Outline

- **minimum cut and maximum flow problems**
- parametric minimum cut problem
- application to proximal mapping

Minimum cut problem

find subset $I \subseteq \{1, 2, \ldots, n\}$ that minimizes

$$C(I) = \sum_{i \in I, j \notin I} A_{ij} + \sum_{i \in I} b_i$$

- $A \in \mathbf{S}^n$ with $A_{ij} = A_{ji} \geq 0$; no sign restrictions on $b \in \mathbf{R}^n$
- cost can be expressed as

$$C(I) = x^T A(\mathbf{1} - x) + b^T x$$

x is the *incidence vector* of I: $x_k = 1$ if $k \in I$, $x_k = 0$ if $k \notin I$

graph interpretation

- optimal two-way partition of n nodes of undirected graph
- first term in $C(I)$ is cost of the *cut* (edges removed by partitioning)

Discrete optimization formulations

binary quadratic maximization

$$\begin{array}{ll} \text{minimize} & -x^T A x + (A\mathbf{1} + b)^T x \\ \text{subject to} & x \in \{0,1\}^n \end{array}$$

cost function is equal to $C(I)$ if x is incidence vector of I

binary piecewise-linear minimization

$$\begin{array}{ll} \text{minimize} & \sum_{i>j} A_{ij} |x_i - x_j| + b^T x \\ \text{subject to} & x \in \{0,1\}^n \end{array}$$

cost function is equal to $C(I)$ if x is incidence vector of I

Convex relaxation

relaxation: replace $x \in \{0,1\}^n$ with $0 \preceq x \preceq \mathbf{1}$ (componentwise)

$$\begin{array}{ll} \text{minimize} & \sum_{i>j} A_{ij}|x_i - x_j| + b^T x \\ \text{subject to} & 0 \preceq x \preceq \mathbf{1} \end{array}$$

we will use LP duality to show that the **relaxation is exact**

- relaxation has an optimal solution $x \in \{0,1\}^n$
- if $x \notin \{0,1\}^n$ is optimal for the relaxation, then rounding x as

$$\hat{x}_i = 1 \quad \text{if } x_i > 1/2, \qquad \hat{x}_i = 0 \quad \text{if } x_i \leq 1/2$$

  gives an integer optimal solution $\hat{x} \in \{0,1\}^n$

Linear program formulation

relaxed problem as LP: introduce matrix variable $Y \in \mathbf{R}^{n\times n}$

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(AY) + b^T x \\ \text{subject to} & Y \geq x\mathbf{1}^T - \mathbf{1}x^T \\ & Y \geq 0 \\ & 0 \leq x \leq \mathbf{1} \end{array}$$

- (componentwise) inequalities on Y are equivalent to

$$Y_{ij} \geq \max\{0, x_i - x_j\}, \quad i, j = 1, \ldots, n$$

- at optimum, $Y_{ij} = \max\{0, x_i - x_j\}$ because $A_{ij} \geq 0$; therefore

$$\mathbf{tr}(AY) + b^T x = \sum_{i>j} A_{ij}|x_i - x_j| + b^T x$$

Dual problem

dual linear program (variables $U \in \mathbf{R}^{n\times n}$, $v, w \in \mathbf{R}^n$)

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T w \\ \text{subject to} & 0 \leq U \leq A \\ & (U - U^T)\mathbf{1} - v + w + b = 0 \\ & v \geq 0, \quad w \geq 0 \end{array}$$

complementary slackness conditions: if x, Y, U, v, w are optimal

$$\begin{aligned} (Y - x\mathbf{1}^T + \mathbf{1}x^T) \circ U &= 0 \\ Y \circ (A - U) &= 0 \\ x \circ v &= 0 \\ (\mathbf{1} - x) \circ w &= 0 \end{aligned}$$

'$\circ$' denotes Hadamard (componentwise) matrix product

Exactness of relaxation

let x be optimal for the relaxation; round x to $\hat{x} \in \{0, 1\}^n$ as

$$\hat{x}_i = 1 \quad \text{if } x_i > 1/2, \qquad \hat{x}_i = 0 \quad \text{if } x_i \leq 1/2$$

- by complementary slackness, if $\hat{x}_i = 1$ and $\hat{x}_j = 0$, then $x_i > x_j$; hence

$$U_{ij} = A_{ij}, \qquad U_{ji} = 0, \qquad v_i = 0, \qquad w_j = 0$$

- implies $\hat{x}^T A(\mathbf{1} - \hat{x}) + b^T \hat{x}$ is equal to the lower bound from relaxation

$$\begin{aligned}
&\hat{x}^T A(\mathbf{1} - \hat{x}) + b^T \hat{x} \\
&\quad = \hat{x}^T (U - U^T)(\mathbf{1} - \hat{x}) - \hat{x}^T v - (\mathbf{1} - \hat{x})^T w + b^T \hat{x} \\
&\quad = \hat{x}^T ((U - U^T)\mathbf{1} - v + w + b) - \hat{x}^T (U - U^T)\hat{x} - \mathbf{1}^T w \\
&\quad = -\mathbf{1}^T w
\end{aligned}$$

Network flow interpretation of dual LP

change of variables (with $b_{+,k} = \max\{b_k, 0\}$, $b_{-,k} = \max\{-b_k, 0\}$)

$$Z = U - U^T, \quad z_{\rm s} = b_- - w, \quad z_{\rm t} = b_+ - v$$

reformulated dual problem

$$\begin{array}{ll} \mbox{maximize} & \mathbf{1}^T z_{\rm s} - \mathbf{1}^T b_- \\ \mbox{subject to} & Z\mathbf{1} - z_{\rm s} + z_{\rm t} = 0 \\ & -A \preceq Z \preceq A, \quad z_{\rm s} \preceq b_-, \quad z_{\rm t} \preceq b_+ \end{array}$$

- $Z_{ij} = -Z_{ji}$ is flow from node i to node j
- $z_{\rm s}$ is vector of flows from an added 'source' node to nodes $1, \ldots, n$
- $z_{\rm t}$ is vector of flows from nodes $1, \ldots, n$ to an added 'sink' node
- maximize flow $\mathbf{1}^T z_{\rm s}$ from source to sink, subject to capacity constraints

exactness of relaxation is known as the **max-flow min-cut theorem**

Outline

- minimum cut and maximum flow problems
- **parametric minimum cut problem**
- application to proximal mapping

Parametric min-cut problem

min-cut problem: take $b = \alpha\mathbf{1} - c$ with α a scalar parameter

$$\begin{array}{ll} \text{minimize} & \sum_{i,j=1}^{n} A_{ij} \max\{0, y_i - y_j\} + (\alpha\mathbf{1} - c)^T y \\ \text{subject to} & y \in \{0,1\}^n \end{array}$$

equivalent LP and its dual

$$\begin{array}{ll} \text{min.} & \mathbf{tr}(AY) + (\alpha\mathbf{1} - c)^T y \\ \text{s.t.} & Y \geq y\mathbf{1}^T - \mathbf{1}y^T \\ & Y \geq 0 \\ & 0 \leq y \leq \mathbf{1} \end{array} \qquad \begin{array}{ll} \text{max.} & -\mathbf{1}^T w \\ \text{s.t.} & 0 \leq U \leq A \\ & (U - U^T)\mathbf{1} - v + w + \alpha\mathbf{1} = c \\ & v \geq 0, \quad w \geq 0 \end{array}$$

primal variables $y \in \mathbf{R}^n$, $Y \in \mathbf{R}^{n \times n}$; dual variables $U \in \mathbf{R}^{n \times n}$, $v, w \in \mathbf{R}^n$

Optimal value function

$$p^\star(\alpha) = \inf_{y \in \{0,1\}^n} \left(\sum_{i,j=1}^{n} A_{ij} \max\{0, y_i - y_j\} + (\alpha \mathbf{1} - c)^T y \right)$$

immediate properties

- $p^\star(\alpha)$ is piecewise-linear and concave
- $p^\star(\alpha) = n\alpha - \mathbf{1}^T c$ for $\alpha \mathbf{1} \le c$, with optimal solution $y = \mathbf{1}$
- $p^\star(\alpha) = 0$ for $\alpha \mathbf{1} \ge c$, with optimal solution $y = 0$

less obvious: $p^\star(\alpha)$ hat at most $n+1$ linear segments

Monotonicity

parametric min-cut problem

$$\begin{array}{ll} \text{minimize} & f_\alpha(y) = y^T A(\mathbf{1} - y) + (\alpha\mathbf{1} - c)^T y \\ \text{subject to} & y \in \{0,1\}^n \end{array}$$

monotonicity of solutions

if y_β is optimal for $\alpha = \beta$ and y_γ is optimal for $\alpha = \gamma > \beta$ then

$$y_\gamma \leq y_\beta$$

- y_γ is zero in all positions where y_β is zero
- implies that optimal value function $p^\star(\alpha)$ has at most $n+1$ segments

proof of monotonicity property

- denote component-wise maximum and minimum of y_β and y_γ as

$$y_{\min} = \min\{y_\beta, y_\gamma\}, \qquad y_{\max} = \max\{y_\beta, y_\gamma\}$$

- (submodularity) it is readily verified that

$$y_{\max}^T A(\mathbf{1} - y_{\max}) + y_{\min}^T A(\mathbf{1} - y_{\min}) \leq y_\beta^T A(\mathbf{1} - y_\beta) + y_\gamma^T A(\mathbf{1} - y_\gamma)$$

- from submodularity and optimality of y_β, y_γ:

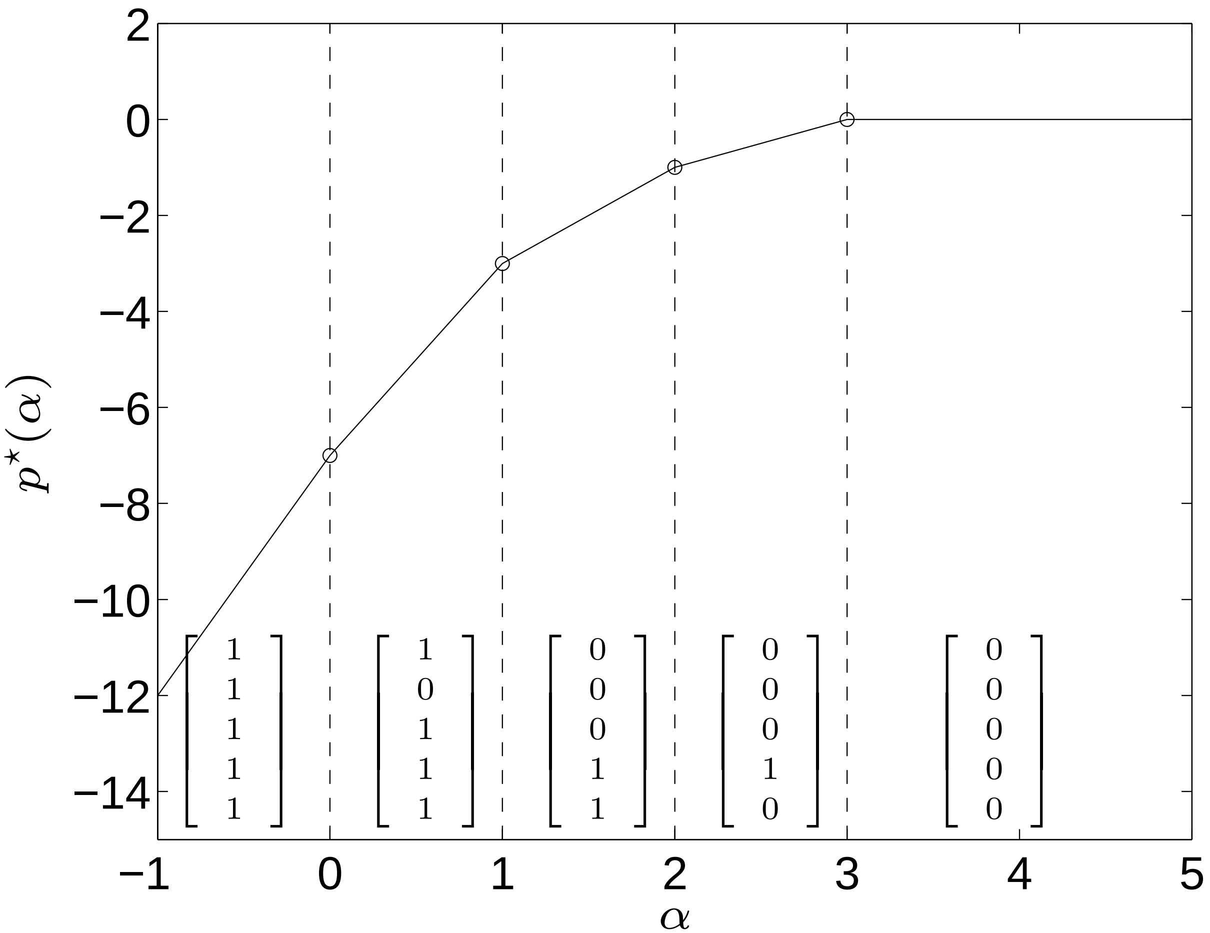
$$\begin{aligned} f_\beta(y_\beta) + f_\beta(y_\gamma) &= f_\beta(y_\beta) + f_\gamma(y_\gamma) - (\gamma - \beta)\mathbf{1}^T y_\gamma \\ &\leq f_\beta(y_{\max}) + f_\gamma(y_{\min}) - (\gamma - \beta)\mathbf{1}^T y_\gamma \\ &= f_\beta(y_{\max}) + f_\beta(y_{\min}) - (\gamma - \beta)\mathbf{1}^T (y_\gamma - y_{\min}) \\ &\leq f_\beta(y_\beta) + f_\beta(y_\gamma) - (\gamma - \beta)\mathbf{1}^T (y_\gamma - y_{\min}) \end{aligned}$$

therefore $\gamma > \beta$ implies $y_\gamma = y_{\min}$

Example

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 1 \\ 3 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$c = \begin{bmatrix} -2 \\ -1 \\ 0 \\ 6 \\ 4 \end{bmatrix}$$



Outline

- minimum cut and maximum flow problems
- parametric minimum cut problem
- **proximal mapping via parametric flow maximization**

Proximal mapping

piecewise-linear convex function

$$h(x) = \sum_{i>j} A_{ij}|x_i - x_j| = \sum_{i,j=1}^{n} A_{ij} \max\{0, x_i - x_j\}$$

proximal mapping: $x = \mathrm{prox}_h(c)$ is the solution of

$$\text{minimize} \quad h(x) + \frac{1}{2}\|x - c\|_2^2$$

- equivalent to a quadratic program
- efficiently computed from solution of parametric minimum cut problem

Quadratic program formulation

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(AY) + \frac{1}{2}\|x - c\|_2^2 \\ \text{subject to} & Y \geq x\mathbf{1}^T - \mathbf{1}x^T \\ & Y \geq 0 \end{array}$$

at optimum, $x = \mathrm{prox}_h(c)$ and $Y_{ij} = \max\{0, y_i - y_j\}$

optimality conditions: there exists a $U \in \mathbf{R}^{n \times n}$ that satisfies

$$0 \leq U \leq A, \qquad x + (U - U^T)\mathbf{1} = c$$

and the complementary slackness conditions

$$(Y - x\mathbf{1}^T + \mathbf{1}x^T) \circ U = 0, \qquad Y \circ (A - U) = 0$$

in particular, if $x_i > x_j$, then $U_{ij} = A_{ij}$ and $U_{ji} = 0$

Relation with parametric min-cut problem

parametric min-cut problem

$$\begin{array}{ll} \text{minimize} & f_\alpha(y) = \sum_{i,j=1}^{n} A_{ij} \max\{0, y_i - y_j\} + (\alpha \mathbf{1} - c)^T y \\ \text{subject to} & y \in \{0,1\}^n \end{array}$$

parametric min-cut solution from proximal mapping

if $x = \mathrm{prox}_h(c)$ then a solution of the parametric min-cut problem is

$$y_i = 1 \quad \text{if } x_i \geq \alpha, \qquad y_i = 0 \quad \text{if } x_i < \alpha$$

proximal mapping from parametric min-cut solution

$$x_i = \sup\{\alpha \mid \text{parametric min-cut problem has a solution with } y_i = 1\}$$

(follows from monotonicity of min-cut solution) proof:

- let $x = \mathrm{prox}_h(c)$ and U the optimal multiplier (page 16); define

$$w = (x - \alpha \mathbf{1})_+, \qquad v = (\alpha \mathbf{1} - x)_+$$

  U, v, w are dual feasible for parametric LP on page 10; therefore

$$f_\alpha(\hat{y}) \geq -\mathbf{1}^T (x - \alpha \mathbf{1})_+ \qquad \forall \hat{y} \in \{0,1\}^n$$

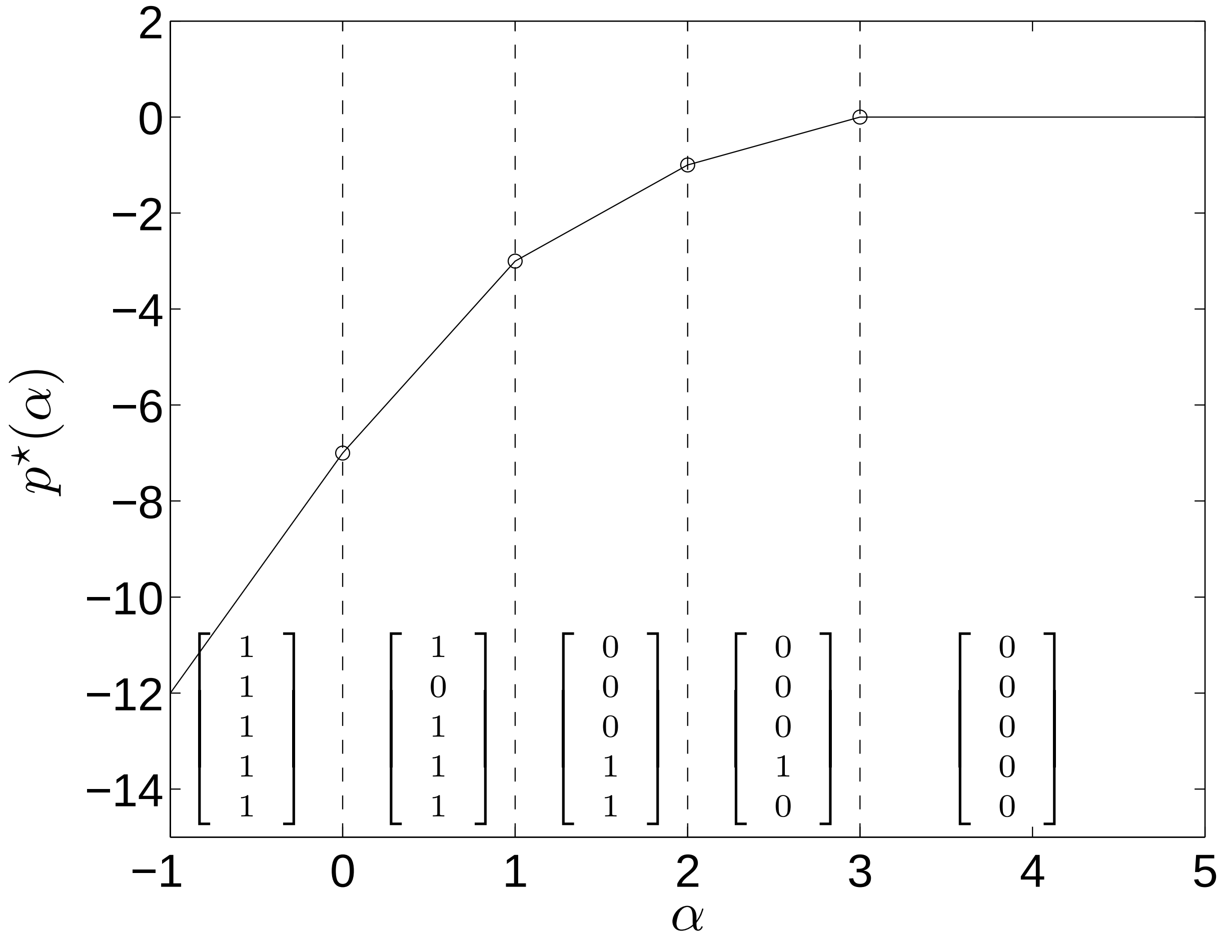
- equality holds for y defined on p. 17:

$$\begin{aligned} f_\alpha(y) &= y^T(U - U^T)(\mathbf{1} - y) + (\alpha \mathbf{1} - c)^T y \\ &= y^T((U - U^T)\mathbf{1} + x - c) - y^T(U - U^T)y - (x - \alpha \mathbf{1})^T y \\ &= -\mathbf{1}^T (x - \alpha \mathbf{1})_+ \end{aligned}$$

  first line follows from $U_{ij} = A_{ij}$, $U_{ji} = 0$ if $y_i = 1$, $y_j = 0$ (see p. 16)

  last line follows from definition of U (p. 16) and construction of y

Example



$$\mathrm{prox}_h(c) = (1, 0, 1, 3, 2)$$

$\mathrm{prox}_h(c)_k$ is value of α at breakpoint where y_k switches from 1 to 0

Summary

proximal mapping of

$$h(x) = \sum_{i>j} A_{ij}|x_i - x_j|$$

can be computed by solving a parametric min-cut/max-flow problem

- very efficiently solved by algorithms from network optimization
- complexity $O(mn\log(n^2/m))$ for general graphs (m is # edges)
- faster algorithms for graphs with special structure

References

- D. Goldfarb and W. Yin, *Parametric maximum flow algorithms for fast total variation minimization*, SIAM Journal on Scientic Computing (2009)

- A. Chambolle and J. Darbon, *On total variation minimization and surface evolution using parametric maximum flows*, International Journal of Computer Vision (2009)

- J. Mairal, R. Jenatton, G. Obozinski, F. Bach, *Network flow algorithms for structured sparsity*, `arxiv.org/abs/1008.5209` (2010)