Supplementary Material: Segmentation with Area Constraints

Marc Niethammer[a,b], Christopher Zach[c]

[a] *Department of Computer Science, University of North Carolina (UNC) at Chapel Hill, USA*
[b] *Biomedical Research Imaging Center, School of Medicine, UNC Chapel Hill, USA*
[c] *Microsoft Research, Cambridge, UK*

Abstract

This document contains supplementary material. In particular, additional segmentation results are shown and validation results are presented. Further, this document contains the full derivations of the prox operators, the ADMM dual energy, and the derivation of the estimate of the relaxed dual energy.

Keywords: Segmentation, area-constraint, branch and bound, alternating direction method of multipliers

To graphically illustrate the behavior of the proposed area-constrained segmentation method Section S.1 shows the segmentation results. Section S.2 presents the detailed validation results of the proposed method in comparison to an unconstrained segmentation, biased normalized cut, normalized cut, seeded watershed, and random walker segmentation. Sections S.3-S.5 contain the derivations for the prox operators, the ADMM dual energy, and the estimate of the relaxed dual energy respectively.

S.1. Full vesicle segmentation results

Figure S1 shows an overview of seed points, the gold standard manual segmentation as well as results for the area-constrained and the unconstrained segmentations for the synaptic vesicle segmentation scenario.

S.2. Statistical significance and detailed segmentation results

Figures S2 and S3 show boxplots for the segmentation results for all the tested methods and statistical significance levels between the methods with respect to mean Dice performance for the synaptic vesicles and the SARS: double-membrane vesicles respectively.

S.3. ADMM prox operators

The prox operator for the update of the consensus variable u is a simple average over the residuals between the local copies and their dual variables:

$$\begin{aligned}
&\operatorname{prox}_{\frac{1}{\sigma}g}(u^A, u^s, \overline{u}^s, \overline{u}^t) = \\
&\operatorname*{argmin}_u \Big(\frac{\sigma}{2} \sum_s (u_s^s - u_s + (z_s^s)^k)^2 + (u_s^A - u_s + (z_s^A)^k) \\
&\quad + \frac{\sigma}{2} \sum_s \Big(\sum_{t:(s,t)\in\mathcal{E}} (\overline{u}_s^s - u_s + (\overline{z}_s^s)^k)^2 \\
&\qquad + \sum_{t:(t,s)\in\mathcal{E}} (\overline{u}_s^t - u_s + (\overline{z}_s^t)^k)^2 \Big) \Big).
\end{aligned}$$

Here, $\mathcal{E}$ is the edge-set. The problem decouples spatially. For any position s on the computational grid we obtain (upon differentiation) the averaging rule of Equation (12).

For the unary term we get

$$\begin{aligned}
&\operatorname{prox}_{\frac{1}{\sigma}f}(q) = \\
&\operatorname*{argmin}_u \sum_s (\rho_s u_s + \imath_{[0,1]}(u_s)) + \frac{\sigma}{2} \sum_s (u_s - q_s)^2.
\end{aligned}$$

This operator decouples spatially also and we therefore need to minimize for every s

$$E(u) = (\rho u + \imath_{[0,1]}(u)) + \frac{\sigma}{2}(u - q)^2.$$

Assume there is no constraint on u, then the minimizer is

$$u = q - \frac{1}{\sigma}\rho,$$

otherwise u will be clamped from which Equa-

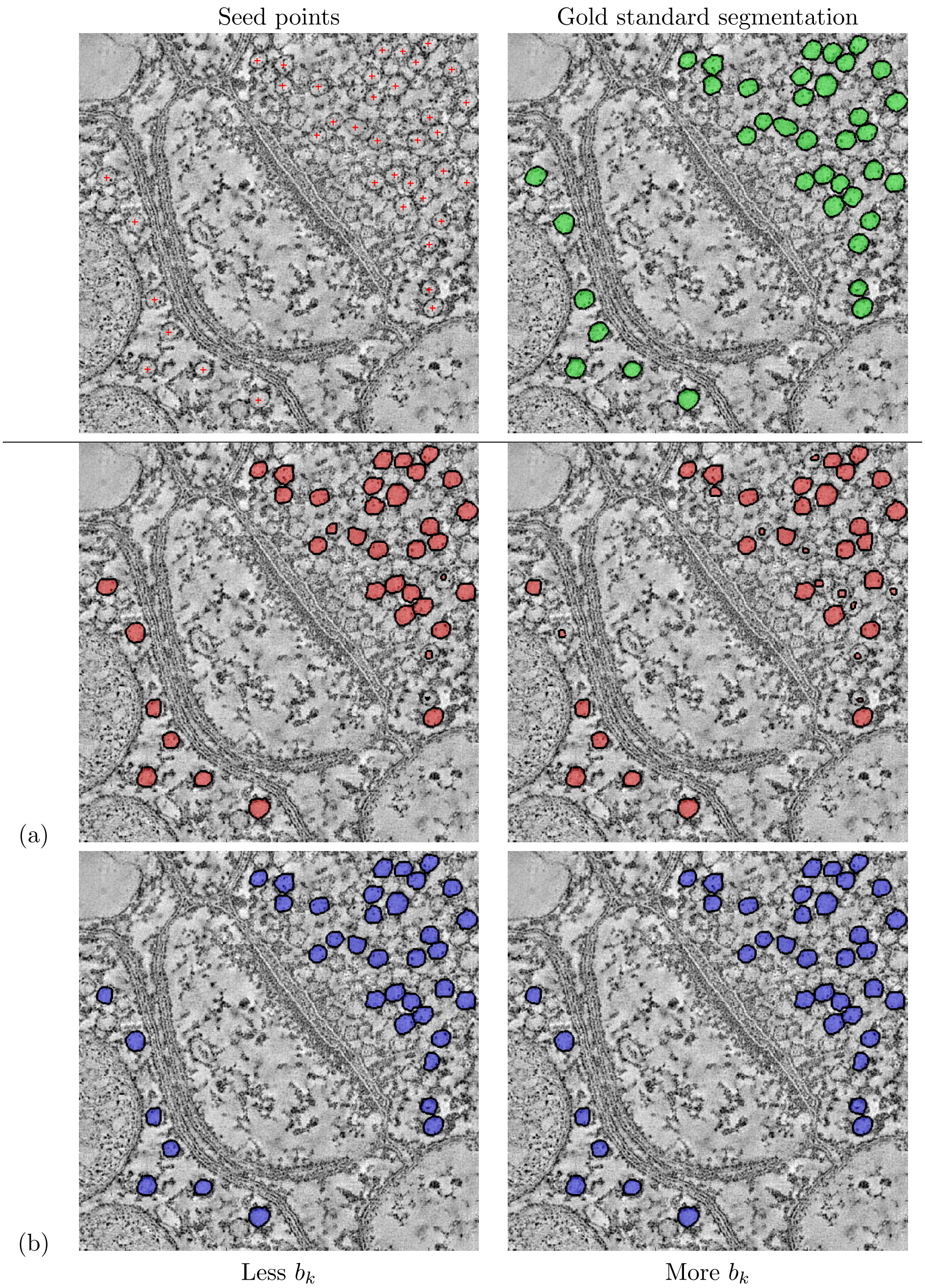
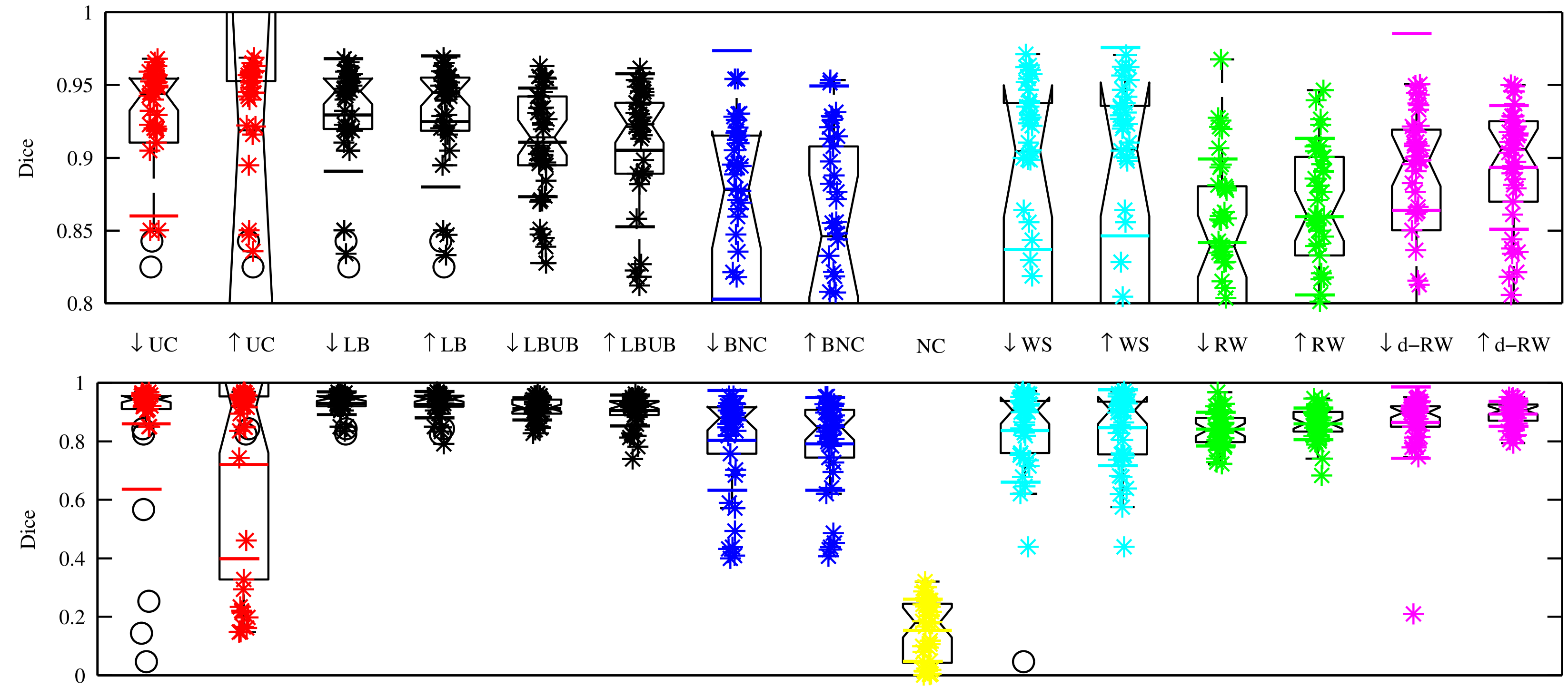
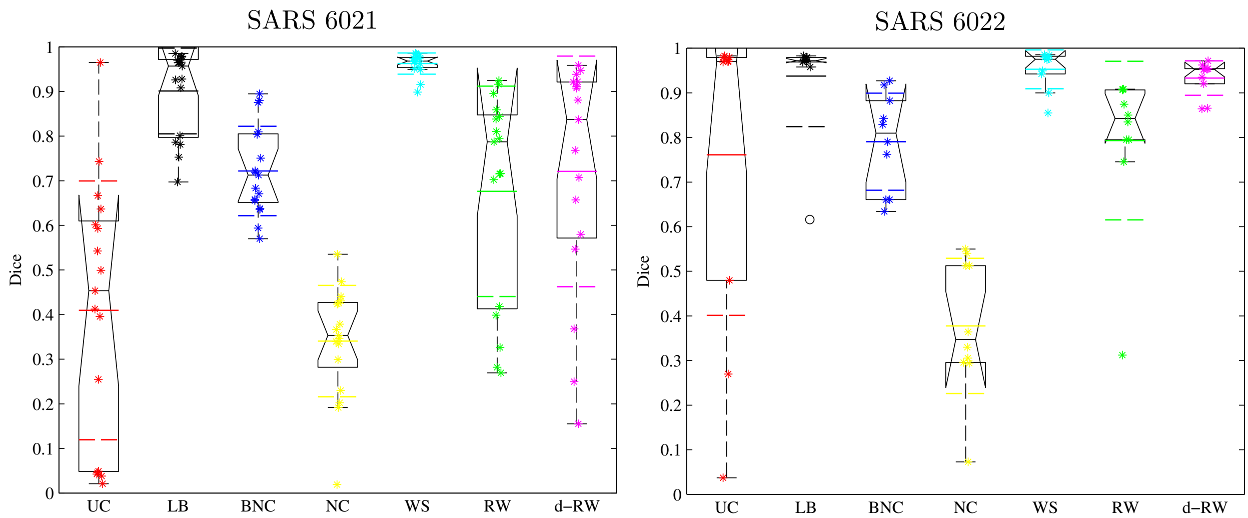


Figure S1: Full vesicle segmentation results: (a) unconstrained, (b) lower bound. Fewer selection regions b_k more (left) and more selection regions b_k (right). The selection regions leave the segmentation results for the area-constrained segmentation largely unaffected, whereas they influence the unconstrained segmentation results greatly. Zoom for best viewing results.



	↑ UC	↓ LB	↑ LB	↓ LBUB	↑ LBUB	↓ BNC	↑ BNC	NC	↓ WS	↑ WS	↓ RW	↑ RW	↓ d-RW	↑ d-RW
↓ UC	0.03	0.07	0.09	0.19	0.24	0.22	0.13	**<1e-4**	0.63	0.76	0.65	0.99	0.93	0.38
↑ UC	-	**1e-4**	**1e-4**	**4e-4**	6e-4	0.17	0.23	**<1e-4**	0.05	0.03	0.03	0.01	0.01	0.001
↓ LB	-	-	0.65	0.03	0.02	**<1e-4**	**<1e-4**	**<1e-4**	**2e-4**	**1e-4**	**<1e-4**	**<1e-4**	**<1e-4**	**3e-4**
↑ LB	-	-	-	0.13	0.08	**<1e-4**	**<1e-4**	**<1e-4**	5e-4	**3e-4**	**<1e-4**	**<1e-4**	**2e-4**	0.002
↓ LBUB	-	-	-	-	0.60	**1e-4**	**<1e-4**	**<1e-4**	0.005	0.003	**<1e-4**	**<1e-4**	0.004	0.06
↑ LBUB	-	-	-	-	-	**3e-4**	**<1e-4**	**<1e-4**	0.02	0.01	**<1e-4**	**3e-4**	0.03	0.28
↓ BNC	-	-	-	-	-	-	0.75	**<1e-4**	0.40	0.21	0.19	0.05	0.08	0.002
↑ BNC	-	-	-	-	-	-	-	**<1e-4**	0.24	0.10	0.07	0.01	0.02	**1e-4**
NC	-	-	-	-	-	-	-	-	**<1e-4**	**<1e-4**	**<1e-4**	**<1e-4**	**<1e-4**	**<1e-4**
↓ WS	-	-	-	-	-	-	-	-	-	0.80	0.89	0.49	0.47	0.05
↑ WS	-	-	-	-	-	-	-	-	-	-	0.84	0.57	0.57	0.03
↓ RW	-	-	-	-	-	-	-	-	-	-	-	0.17	0.36	**<1e-4**
↑ RW	-	-	-	-	-	-	-	-	-	-	-	-	0.89	0.003
↓ d-RW	-	-	-	-	-	-	-	-	-	-	-	-	-	0.15

Figure S2: Statistical significance (p-value) for the null-hypothesis that there are no segmentation differences as measured by the mean Dice similarity coefficients. Results are for vesicle segmentation with small (↓) and larger (↑) number of selection areas, $\mathcal{B}_k$. Unconstrained (UC) segmentation, lower (LB), and lower and upper bound (LBUB) area-constrained segmentations. Biased normalized cut (BNC), normalized cut (NC), seeded watershed (WS), random walker (RW) and random walker with default settings (d-RW). Bold: best results. P-values between Dice results from permutation test (for 100,000 permutations each). Bold: statistically significant at significance level $\alpha = 0.05$ after Bonferroni correction (for 105 tests performed in total). Middle: overview of the individual Dice coefficients for the different segmentation methods. Top: zoom-in to Dice between 0.8 and 1.0. Solid lines indicate mean plus/minus one standard deviation. Circles indicate potential outliers. Box center indicates the median and the box extends between the 25th and the 75th percentile of the data.



	LB	BNC	NC	WS	RW	d-RW
UC	**<1e-4**	**1e-4**	0.38	**<1e-4**	0.01	0.003
LB	-	**1e-4**	**<1e-4**	0.01	**3e-4**	0.008
BNC	-	-	**<1e-4**	**<1e-4**	0.47	0.99
NC	-	-	-	**<1e-4**	**1e-4**	**<1e-4**
WS	-	-	-	-	**<1e-4**	**<1e-4**
RW	-	-	-	-	-	0.60

	LB	BNC	NC	WS	RW	d-RW
UC	0.17	0.81	0.008	0.19	0.80	0.21
LB	-	0.01	**<1e-4**	0.89	0.02	0.98
BNC	-	-	**<1e-4**	**3e-4**	0.98	**5e-4**
NC	-	-	-	**<1e-4**	**1e-4**	**<1e-4**
WS	-	-	-	-	**3e-4**	0.30
RW	-	-	-	-	-	**8e-4**

Figure S3: Statistical significance (p-value) for the null-hypothesis that there are no segmentation differences as measured by the mean Dice similarity coefficients. Unconstrained (UC) segmentation, lower (LB), and lower and upper bound (LBUB) area-constrained segmentations. Biased normalized cut (BNC), normalized cut (NC), seeded watershed (WS), random walker (RW) and random walker with default settings (d-RW). Bold: best results. P-values between Dice results from permutation test (for 100,000 permutations each). Bold: statistically significant at significance level $\alpha = 0.05$ after Bonferroni correction (for 21 tests performed in total for each image). Top: overview of the individual Dice coefficients for the different segmentation methods. Solid lines indicate mean plus/minus one standard deviation. Circles indicate potential outliers. Box center indicates the median and the box extends between the 25th and the 75th percentile of the data.

tion (13) follows. For the area-based term, we have

$$\operatorname{prox}_{\frac{1}{\sigma} f_A}(q) = \operatorname*{argmin}_{u^A} \imath\left\{A_l \le \sum_s u_s^A \le A_u\right\} + \frac{\sigma}{2}\sum_s (u_s^A - q_s)^2.$$

If the area is too small, all values are uniformly increased, if it is too large they are uniformly decreased. Consider the case $\sum_s u_s < A_l$, then the projection will result in $\sum_s u_s = A_l$. Adding this constraint through a Lagrangian multiplier λ_e results in

$$E(q) = \frac{\sigma}{2}\sum_s (u_s^A - q_s)^2 - \lambda_e(\sum_s u_s^A - A_l).$$

At a critical point the KKT conditions

$$\begin{aligned} \sigma(u_s - q_s) - \lambda_e &= 0, \\ \sum_s u_s &= A_l, \end{aligned}$$

from which follows that the values are uniformly raised (by addition of λ_e/σ) so that the area constraint is fulfilled. The same reasoning holds for the upper bound. Hence Equation (14) follows. Since the edge-terms do not decouple we need to compute the joint prox operator

$$\operatorname{prox}_{\frac{1}{\sigma} f_{st}}(u, v) = \operatorname*{argmin}_{\overline{u}^s, \overline{u}^t} \sum_{(s,t)} c_{st}|\overline{u}_s^s - \overline{u}_t^t| + \frac{\sigma}{2}\sum_{(s,t)} (\overline{u}_s^s - u)^2 + (\overline{u}_t^t - v)^2,$$

which decouples for every edge pairing[1]. Therefore, for every edge pairing

$$\operatorname{prox}_{\frac{1}{\sigma} f_{st(s,t)}}(u, v) = \operatorname*{argmin}_{\overline{u}_s^s, \overline{u}_t^t} c_{st}|\overline{u}_s^s - \overline{u}_t^t| + \frac{\sigma}{2}\left((\overline{u}_s^s - u)^2 + (\overline{u}_t^t - v)^2\right).$$

Considering the three cases (1): $\overline{u}_s^s = \overline{u}_t^t$, (2): $\overline{u}_s^s < \overline{u}_t^t$, (3) $\overline{u}_s^s > \overline{u}_t^t$, we get in the first case

$$\operatorname{argmin} E(\overline{u}_s^s) = \operatorname{argmin} \frac{\sigma}{2}\left((\overline{u}_s^s - u)^2 + (\overline{u}_s^s - v)^2\right) = \frac{u+v}{2}.$$

[1]Note that this works, because the edges are chosen to conform with right-sided differences if working on a grid (or the equivalent for a general graph). Thus $\overline{u}_s^s$ denotes an outgoing edge and $\overline{u}_t^t$ an incoming edge.

In the second case,

$$\begin{aligned} &\operatorname*{argmin}_{\overline{u}_s^s, \overline{u}_t^t} c_{st}(\overline{u}_t^t - \overline{u}_s^s) + \frac{\sigma}{2}\left((\overline{u}_s^s - u)^2 + (\overline{u}_t^t - v)^2\right) \\ &\qquad = \begin{pmatrix} u + \frac{c_{st}}{\sigma} & v - \frac{c_{st}}{\sigma} \end{pmatrix} \end{aligned}$$

and similarly in the third case

$$\begin{aligned} &\operatorname*{argmin}_{\overline{u}_s^s, \overline{u}_t^t} c_{st}(\overline{u}_s^s - \overline{u}_t^t) + \frac{\sigma}{2}\left((\overline{u}_s^s - u)^2 + (\overline{u}_t^t - v)^2\right) \\ &\qquad = \begin{pmatrix} u - \frac{c_{st}}{\sigma} & v + \frac{c_{st}}{\sigma} \end{pmatrix}, \end{aligned}$$

which yields the prox operator of Equation (15). This leaves the prox operator for g. For a point $s \notin \mathcal{S} \cup \mathcal{T}$

$$\begin{aligned} \operatorname{prox}_{\frac{1}{n\sigma} g_s}(u) &= \operatorname*{argmin}_{u_s} \imath_{[0,1]}(u_s) + \frac{n\sigma}{2}(u_s - u)^2 \\ &= \min(1, \max(0, u)) \end{aligned}$$

For points s in $\mathcal{S}$ or $\mathcal{T}$ the prox operator simply returns 1 or 0 respectively and immediately yields the prox operator of Equation (16).

S.4. ADMM dual energy

The z variables correspond to the normalized dual variables ($p = \sigma z$). To compute the dual energy requires the computations of the conjugate functions for f_s, f_{st}, and f_A. We obtain

$$f_s^*(p) = \sup_{u \in [0,1]} (up - \rho u) = [p - \rho]_+.$$

For the area-based term

$$\begin{aligned} f_A^*(p) &= \sup_u \{p^T u - \imath\{A_l \le \sum_s A_s u_s \le A_u\}\} \\ &= \sup_{u: A_l \le \sum A_s u_s \le A_u} (p^T u) \\ &= \begin{cases} A_u \max_s \frac{p_s}{A_s}, & \text{if } \exists s: p_s \ge 0, \\ A_l \max_s \frac{p_s}{A_s}, & \text{otherwise.} \end{cases} \end{aligned}$$

For the edge-term

$$f_{st}^*(p_s, p_t) = \sup_{u,v} \left[p_s u + p_t v - c_{st}|u - v|\right],$$

which is equal the maximum over all three cases $u < v$, $u = v$, $u > v$. For $u = v$

$$\sup_u \left[p_s u + p_t u\right] = \imath\{p_s + p_t = 0\}.$$

Hence $p_s + p_t = 0$ for $f^*_{st} < \infty$. Testing the other two cases under this restriction yields

$$\begin{aligned}\sup_{u>v}\ [(p_s - c_{st})(u-v)] &= \imath\{p_s \le c_{st}\},\\ \sup_{u<v}\ [(p_s + c_{st})(u-v)] &= \imath\{p_s \ge -c_{st}\}.\end{aligned}$$

Therefore

$$f^*_{st}(p_s, p_t) = \imath\{p_s + p_t = 0 \wedge |p_s| \le c_{st}\}.$$

The overall primal energy can be written abstractly as

$$\begin{aligned}E(\mathcal{U}, \mathcal{V}_k) &= \sum_k h_k(\mathcal{V}_k) + \sum_k \imath\{\mathcal{U} = \mathcal{V}_k\}\\ &= \sum_k h_k(\mathcal{V}_k) + \imath\{A\overline{\mathcal{U}} = 0\},\end{aligned}$$

s.t. $\mathcal{U} \in [0,1]$, $\mathcal{U}_s = 1$ for $s \in \mathcal{S}$; $\mathcal{U}_s = 0$ for $s \in \mathcal{T}$,

where $\overline{\mathcal{U}} = (\mathcal{U}, \dots, \mathcal{V}_k, \dots)^T$ and

$$A = \begin{pmatrix} -I & I & & \\ -I & & I & \\ \vdots & & & \ddots & \\ -I & & & & I \end{pmatrix}.$$

The conjugate function for $f_c = \imath\{A\overline{\mathcal{U}} = 0\}$ is

$$\begin{aligned}f^*_c(\overline{\mathcal{P}}) &= \sup_{\overline{\mathcal{U}}} \left[\overline{\mathcal{P}}^T\overline{\mathcal{U}} - \imath\{A\overline{\mathcal{U}} = 0\}\right]\\ &= \begin{cases} 0, & \text{for } s \in \mathcal{T},\\ \mathcal{P} + \sum_k Q_k, & \text{for } s \in \mathcal{S},\\ \sup\limits_{\mathcal{U}\in[0,1]} \left[\mathcal{P}\mathcal{U} + \sum_k Q_k \mathcal{U}\right], & \text{otherwise},\end{cases}\\ &= \begin{cases} 0, & \text{for } s \in \mathcal{T},\\ \mathcal{P} + \sum_k Q_k, & \text{for } s \in \mathcal{S},\\ [\mathcal{P} + \sum_k Q_k]_+, & \text{otherwise},\end{cases}\end{aligned}$$

where $\overline{\mathcal{P}} = (\mathcal{P}, \dots, \mathcal{Q}_k, \dots)^T$ and the derivations are based on the fact that the supremum is only achieved (in all cases) if $\mathcal{U} = \mathcal{V}_k$. The results hold pointwise. Note that we are free to choose a $\mathcal{P}$. Note that for $s \notin \mathcal{S} \cup \mathcal{T}$

$$\begin{aligned}&\left(\sum_k h_k(\mathcal{V}_k)\right)^* =\\ &\sup_{\mathcal{U}\in[0,1], \mathcal{V}_k} \sum_k (\mathcal{Q}_k^T \mathcal{V}_k - h_k(\mathcal{V}_k)) + \mathcal{P}^T\mathcal{U}\\ &= \sum_k h^*_k(\mathcal{Q}_k) + \sup_{\mathcal{U}\in[0,1]} \mathcal{P}^T\mathcal{U} = \sum_k h^*_k(\mathcal{Q}_k) + \sum_s [\mathcal{P}_s]_+.\end{aligned}$$

Overall, we obtain

$$\left(\sum_k h_k(\mathcal{V}_k)\right)^* = \sum_k h^*_k(\mathcal{Q}_k) + \sum_{s\in\mathcal{S}} \mathcal{P}_s + \sum_{s\notin\mathcal{T}\cup\mathcal{S}} [\mathcal{P}_s]_+.$$

The dual program to

$$\inf_y \ \{f(y) + g(y)\}$$

is according to Fenchel duality

$$\sup_z \ \{-f^*(z) - g^*(-z)\}.$$

Therefore the overall dual is

$$\begin{aligned}(E(\mathcal{U}, \mathcal{V}_k))^* = -\Big(&\sum_k h^*_k(\mathcal{Q}_k) + \sum_{s\in\mathcal{S}} \left(-\sum_k (Q_k)_s\right)\\ &+ \sum_{s\notin\mathcal{T}\cup\mathcal{S}} \left([\mathcal{P}_s]_+ + [-\mathcal{P}_s - \sum_k (Q_k)_s]_+\right)\Big).\end{aligned}$$

For the ADMM this results in the dual energy

$$\begin{aligned}&E^*(z^s, z^A, \overline{z}^s, \overline{z}^t) =\\ &-\Bigg\{\sum_s f^*_s(z^s_s) + f^*_A(z^A) + \sum_{(s,t)} f^*_{st}(\overline{z}^s_s, \overline{z}^t_t)\\ &\quad + \sum_{s\in\mathcal{S}} (-z^s_s - z^A_s - \overline{z}^s_s - \overline{z}^t_s)\\ &+ \sum_{s\notin\mathcal{T}\cup\mathcal{S}} ([z_s]_+ + [-z_s - z^s_s - z^A_s - \overline{z}^s_s - \overline{z}^t_s]_+)\Bigg\}.\end{aligned}$$

S.5. Estimate of the relaxed dual energy

A lower bound for the dual energy can be obtained by adjusting the dual variables for the terms which would otherwise lead to a $-\infty$ estimate before convergence. For the edge-based variables we therefore need to find a dual variable pair $(\tilde{p}_s, \tilde{p}_t)$ which is as close as possible to the current estimate (p_s, p_t) while fulfilling the constraint on this edge variable, i.e., we need to solve

$$\operatorname*{argmin}_{u,v} \imath\{u+v = 0, |u| \le c_{st}\} + \frac{1}{2}\left((u-p_s)^2 + (v-p_t)^2\right),$$

which is equivalent for $v = -u$ to

$$\operatorname*{argmin}_{u} \imath\{|u| \le c_{st}\} + \frac{1}{2}\left((u-p_s)^2 + (-u-p_t)^2\right).$$

In the unconstrained case the solution is

$$u = \frac{p_s - p_t}{2}, \quad v = -u,$$

therefore the projection is

$$\Pi(p_s, p_t) = \begin{cases} (c, -c), & \text{for } p_s - p_t > 2c_{st}, \\ (-c, c), & \text{for } p_s - p_t < -2c_{st}, \\ (\frac{p_s - p_t}{2}, \frac{p_t - p_s}{2}), & \text{otherwise.} \end{cases}$$

Since within the solution process the dual variable z is nowhere computed, we are free to choose it as desired. With the objective being a lower bound which is as large as possible this amounts to finding (at every point s) z such that

$$z = \operatorname*{argmin}_x \, [x]_+ + [-x + p]_+$$

where $p \in \mathbb{R}$ is the negative sum of the dual variables (z_s^s, z_s^A, $\overline{z}_s^s$, $\overline{z}_s^t$). There will be an infinite number of solutions, but we get

$$\min_x \, [x]_+ + [-x + p]_+ = [p]_+.$$

The overall (finite-valued) dual energy which can be used as the current lower bound also before full convergence of the iterations is therefore

$$\begin{aligned} &E^*(z^s, z^A, \overline{z}^s, \overline{z}^t) = \\ &-\Bigg\{ \sum_s f_s^*(z_s^s) + f_A^*(z^A) + \sum_{(s,t)} f_{st}^*((\overline{z}_s^s)^\Pi, (\overline{z}_t^t)^\Pi) \\ &\qquad + \sum_{s \in \mathcal{S}} (-z_s^s - z_s^A - (\overline{z}_s^s)^\Pi - (\overline{z}_s^t)^\Pi) \\ &\qquad + \sum_{s \notin \mathcal{T} \cup \mathcal{S}} ([-z_s^s - z_s^A - (\overline{z}_s^s)^\Pi - (\overline{z}_s^t)^\Pi]_+) \Bigg\}, \end{aligned}$$

where $(\overline{z}_s^s)^\Pi$ and $(\overline{z}_s^t)^\Pi$ denote the projected edge variables.