Optimization of Graph Total Variation via Active-Set-based Combinatorial Reconditioning
— Supplementary Material —

Zhenzhang Ye
TU Munich
zhenzhang.ye@tum.de

Thomas Möllenhoff
TU Munich
thomas.moellenhoff@tum.de

Tao Wu
TU Munich
tao.wu@tum.de

Daniel Cremers
TU Munich
cremers@tum.de

Lemma 1. *Let h be C^2 with $l_h I \preceq \nabla^2 h(\cdot) \preceq L_h I$ for some constants l_h, $L_h > 0$. Then the gradient descent on $\min_x \; h(Ax+b)$ with step size $1/t = 2/(L_h \sigma_{max}(A)^2 + l_h \sigma_{min>0}(A)^2)$ satisfies*

$$\|x^{k+1} - x^*\| \leq \frac{\varphi - 1}{\varphi + 1} \|x^k - x^*\|, \tag{1}$$

with $\varphi = \kappa(A)^2 \cdot \kappa(h)$, $\kappa(h) := L_h / l_h$.

Proof of Lemma 1. Clearly $t > (L_h \cdot \lambda_{\max}(A^\top A))/2$, so classical theory guarantees $x^k \to x^*$. First note that

$$x^{k+1} - x^k = -\frac{1}{t} A^\top \nabla h(Ax^k + b) \in \operatorname{ran} A^\top, \tag{2}$$

and consequently $x^k - x^0 \in (\ker A)^\perp$, also $x^* - x^k \in (\ker A)^\perp$. Inserting the gradient step for x^{k+1} yields

$$\|x^{k+1} - x^*\| = \|x^k - x^* - \frac{1}{t} A^\top (\nabla h(Ax^k + b) - \nabla h(Ax^* + b))\|. \tag{3}$$

From the mean value theorem it follows

$$\nabla h(Ax^k + b) - \nabla h(Ax^* + b) = M(Ax^k - Ax^*), \tag{4}$$

with $M = \int_0^1 \nabla^2 h(Ax^* + b + \alpha(Ax^* - Ax^k)) \mathrm{d}\alpha$. Since $l_h I \preceq \nabla^2 h(\cdot) \preceq L_h I$ we have $l_h I \preceq M \preceq L_h I$. This yields due to $x^k - x^* \in (\ker A)^\perp$ that

$$\|x^{k+1} - x^*\| = \|(I - \frac{1}{t} A^\top M A)(x^k - x^*)\| \leq \max\{|1 - l_h \sigma_{\min>0}(A)^2/t|, |1 - L_h \sigma_{\max}(A)^2/t|\} \cdot \|x^k - x^*\|. \tag{5}$$

The choice $t = (l_h \sigma_{\min>0}(A)^2 + L_h \sigma_{\max}(A)^2)/2$ minimizes the above rate and yields the desired result. ☐