

## SUPPLEMENTARY MATERIALS: Nonparametric posterior learning for emission tomography with multimodal data\*

Fedor Goncharov<sup>†</sup>, Éric Barat<sup>†</sup>, and Thomas Dautremer<sup>†</sup>

**SM1. Construction of the common probability space.** Let  $(\Omega', \mathcal{F}', P')$  be the probability space on which the stationary spatio-temporal Poisson point process  $Z^t$  is defined ( $Z^t$  has values in  $Z \times [0, +\infty)$ ;  $Z$  being the space of LORs). Sinogram  $Y^t$  is obtained from binning  $Z^t$  to detector elements, therefore process  $Y^t$  is a well-defined random variable on  $(\Omega', \mathcal{F}', P')$ . Measure-theoretic construction of  $Z^t$  and  $(\Omega', \mathcal{F}', P')$  can be found, for example, in [SM5], Section 9.2, Example 9.2(b).

Algorithms 3 and 4 rely only on perturbed intensities  $\tilde{\Lambda}_{\mathcal{M}}^t$  and  $\tilde{\Lambda}_b^t$  for which we show that they can be expressed as functions of random weighting of the list-mode data:

$$G^t = \{\delta_{(k,i)} : (k,i) - k^{\text{th}} \text{ photon was detected at detector } i\},$$

where  $\delta_{(k,i)} \in \{0, 1\}$ . Indeed, from Step 4 in Algorithm 3 we can see that  $\tilde{\Lambda}_{\mathcal{M}}^t$  is a function of  $\tilde{\Lambda}^t$  for which the following representation holds

$$(SM1.1) \quad \tilde{\Lambda}_i^t = t^{-1} \sum_{k=1}^{N^t} \delta_{(k,i)} \tilde{w}_k, \quad i \in \{1, \dots, d\},$$

$$(SM1.2) \quad \{\tilde{w}_k\}_{k=1}^{N^t} \stackrel{iid}{\sim} \Gamma(1, 1),$$

where  $N^t$  is the total number of photons. For  $\tilde{\Lambda}_b^t$  in Step 5 of Algorithm 4 we have the following representation:

$$(SM1.3) \quad \tilde{\Lambda}_{b,i}^t | \tilde{\Lambda}_{\mathcal{M},i}^t = (\theta^t + t)^{-1} \left( \sum_{k=1}^{N^t} \delta_{(k,i)} w_k + w_{\mathcal{M},i} \right),$$

$$(SM1.4) \quad \{w_k\}_{k=1}^{N^t}, w_{\mathcal{M},i} \sim \Gamma(\theta^t \tilde{\Lambda}_{\mathcal{M},i}^t, 1).$$

Note that  $w_{\mathcal{M},i}$  can be constructed simply from vector  $\omega_i \stackrel{iid}{\sim} \Gamma(1, 1)$  which are also independent of all other variables. Therefore, from formulas (SM1.1)–(SM1.4) one can see that perturbations  $\tilde{\Lambda}_{\mathcal{M}}^t$  and  $\tilde{\Lambda}_b^t$  depend on data  $Y^t$  and on family of mutually independent weights  $(\{(w_k, \tilde{w}_k)\}_{k=1}^{\infty}, \{\omega_i\}_{i=1}^d)$  which are also independent of  $Y^t$ . Therefore, the common probability space can be defined as follows:

$$(SM1.5) \quad (\Omega, \mathcal{F}, P) = (\Omega' \times \Omega_w \times \Omega_{\tilde{w}} \times \Omega_\omega, \mathcal{F}' \times \mathcal{F}_w \times \mathcal{F}_{\tilde{w}} \times \mathcal{F}_\omega, P' \times P_w \times P_{\tilde{w}} \times P_\omega),$$

\*Submitted to the editors October 26, 2022.

**Funding:** This work is partly supported by the 'MMIPROB' project funded by ITMO Cancer (France).

<sup>†</sup>Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France ([fedor.goncharov@cea.fr](mailto:fedor.goncharov@cea.fr)).

where  $(\Omega_w, \mathcal{F}_w, P_w)$ ,  $(\Omega_{\tilde{w}}, \mathcal{F}_{\tilde{w}}, P_{\tilde{w}})$ ,  $(\Omega_\omega, \mathcal{F}_\omega, P_\omega)$  are the probability spaces for infinite sequences of i.i.d r.v.s  $\{w_k\}_{k=1}^\infty$ ,  $\{\tilde{w}_k\}_{k=1}^\infty$ ,  $w_k \sim \Gamma(1, 1)$ ,  $\tilde{w}_k \sim \Gamma(1, 1)$  and for  $\{\omega_i\}_{i=1}^d \sim \Gamma(1, 1)$ , respectively. This construction originates to [SM22] and similar ones have been also used in [SM23].

**SM2. Limit theorems for stationary Poisson processes.** Let

$$(SM2.1) \quad Y^t \sim \text{Po}(\Lambda t), \Lambda > 0, t \in [0, +\infty).$$

The following classical result is a composition of theorems 9.3, 4.1 and 7.5 (pp. 306, 350, 417, respectively) from [SM13].

**Theorem SM2.1.** *Let  $\{Y^t\}$ ,  $t \in (0, +\infty)$  be the Poisson process defined in (SM2.1). Then,*

(i)

$$(SM2.2) \quad \frac{Y^t}{t} \xrightarrow{a.s.} \Lambda \text{ when } t \rightarrow +\infty.$$

(ii)

$$(SM2.3) \quad \frac{Y^t - \Lambda t}{\sqrt{\Lambda t}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ when } t \rightarrow +\infty.$$

(iii)

$$(SM2.4) \quad \liminf_{t \rightarrow +\infty} (\limsup_{t \rightarrow +\infty}) \frac{Y^t - \Lambda t}{\sqrt{\Lambda t \log \log t}} = -\sqrt{2}(\sqrt{2}) \text{ a.s.},$$

where  $\xrightarrow{a.s.}$ ,  $\xrightarrow{d}$  denote the convergence almost surely and in distribution, respectively, *a.s.* denotes that statement holds for almost any trajectory  $Y^t$ ,  $t \in (0, +\infty)$ .

**SM3. A remark on one recent bootstrap algorithm for ET.** A very recent and similar to ours sampling algorithm was proposed in [SM9] provided with a very extensive experiment both on synthetic and real PET-MRI data. The algorithm there is also of bootstrap-type, it is based on optimization of a randomized KL-distance and in fact, it coincides up to minor details with Algorithm 4 for  $\theta^t \equiv 0$ . The difference is that data  $\mathcal{M}$  are used there to construct very special penalty  $\varphi(\lambda) = \varphi_{\mathcal{M}}(\lambda)$  of Bowsher type (see the discussion in subsection 2.4). This penalty satisfies the assumptions in (2.11) and (2.12), so our Theorems 6.5 and 6.10 serve as a theoretical foundation for the algorithms in [SM9]. A nice practical feature of our approach is that  $\theta^t$  has clear physical interpretation of the effect of MRI data on samples (see Remark 4.3), whereas large number of parameters in Bowsher-type penalties have no such easy interpretations making the problem of (posterior) calibration cumbersome.

The aforementioned minor differences between algorithms consist in the way  $Y^t$  (in [SM9]) are stochastically perturbed. From the first look this seems to be only a technical question, however, we think that it is not. From the derivation of NPL in section 4 one can see that uncertainty propagates via the KL-projection in (4.5). Moreover, we retrieve version of WLB of [SM22] adapted for ET as a particular case of Algorithm 4 when choosing the scale parameter  $\theta^t = 0$  in the nonparametric prior in (4.11). This is fully coherent with

the derivation of NPL in [SM20] and nonparametric posterior bootstrap with MDP-prior in [SM10], where the classical WLB algorithm from [SM22] is retrieved back as a particular case when choosing the concentration parameter  $\alpha = 0$  ( $c = 0$  in [SM10]). On the other hand, the derivation in [SM9] strongly relies on model with finite data and it is claimed that the resulting algorithm is also a version of WLB, however, in this case for us is not clear which randomized functional stands behind this procedure as it was with KL-distance in our work.

**SM4. Practical interpretation of slow mixing in MCMC.** In practice produced samples by the Markov chain are used to compute credible intervals for weighted means in certain subregions of reconstructed images. Let  $h \in \mathbb{R}^p$  be a weighting mask which corresponds to subregion  $\Omega \subset \{1, \dots, p\}$ . For example, if  $h_j = \frac{1}{\#\Omega}$  for pixel  $j \in \Omega$  and  $h_j = 0$  otherwise, then  $h^T \lambda$  gives the average tracer concentration in subregion  $\Omega$ . Let  $N$  be the number of generated samples which we denote by  $\{\lambda_k^t\}_{k=1}^N$ . Then, the posterior mean of  $h^T \lambda$  can be approximated by the following expression:

$$(SM4.1) \quad \widehat{f}_{h,N}^t = \frac{1}{N} \sum_{k=1}^N h^T \lambda_k^t,$$

The variance of estimator  $\widehat{f}_{h,N}^t$  can be approximated as follows:

$$(SM4.2) \quad \begin{aligned} \text{var}(\widehat{f}_{h,N}^t | Y^t, t) &= \frac{1}{N^2} \sum_{k=1}^N \sum_{s=1}^N \text{cov}(h(\lambda_k^t), h(\lambda_s^t) | Y^t, t) \\ &\asymp \frac{\sigma^2}{N} \left(1 + 2 \sum_{k=1}^{\infty} \rho_k^t(h)\right), \end{aligned}$$

where

$$(SM4.3) \quad \rho_k^t(h) = \text{corr}(h^T \lambda_1^t, h^T \lambda_{k+1}^t | Y^t, t), \quad \sigma^2 = \text{var}(h^T \lambda).$$

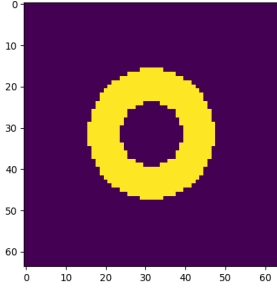
In [SM18] it was shown, in particular, that  $\rho_k^t(h) \asymp (\gamma^t(h))^k$ , so from above formulas we get the following expression for the variance of  $\widehat{f}_{h,N}^t$  (modulo a universal multiplicative factor):

$$(SM4.4) \quad \text{var}(\widehat{f}_{h,N}^t | Y^t, t) \asymp \frac{\sigma^2}{N} \left( \frac{1 + \gamma^t(h)}{1 - \gamma^t(h)} \right) \approx \frac{\sigma^2}{N} \left( \frac{1 + \gamma(h)}{1 - \gamma(h)} \right),$$

where  $\gamma^t(h)$ ,  $\gamma(h)$  are defined in (3.5) and (3.7), respectively. The rule of thumb in [SM2] tells to choose  $N$  such that empirical variance of  $\widehat{f}_{h,N}^t$  does not exceed 1% of  $\sigma^2$ , which is then translated to the following rule:

$$(SM4.5) \quad \frac{\text{var}(\widehat{f}_{h,N}^t | Y^t, t)}{\sigma^2} < 0.01 \Rightarrow N \gtrsim 100 \times \left( \frac{1 + \gamma(h)}{1 - \gamma(h)} \right) \approx +\infty \text{ for } h = h_m, m \gg 1.$$

Therefore, to estimate reliably the average signal using mask  $h \in \mathbb{R}^p$ , one needs almost infinite number of samples if  $h$  contains a high-frequency component in terms of basis  $\{h_k\}_{k=1}^p$ .



$\lambda_*$  – image of size  $64 \times 64$  (see Figure SM1),  
 $A$  – Radon transform matrix of size  $4096 \times 4096$ ,  
 prior  $\pi_j = \Gamma(1, 1)$ ,  
 time  $t = 10^2, 10^{10}$  ( $\sim$  photons per LOR),  
 initial point:  $\lambda_*$ ,  
 burn-in samples: 1000,  
 number of samples for the output: 2000

Figure SM1:  $\lambda_*$

**SM5. Numerical experiment for the Gibbs sampler.**<sup>1</sup> According to (3.8) we choose  $\lambda_* \succ 0$ , where  $\lambda_*|_{\text{ring}} = 2.0$ ,  $\lambda_*|_{\text{background}} = 1.0$ , where radius of the inner circle  $r_{\text{in}} = 0.25$  and of the outer  $r_{\text{out}} = 0.5$ , and the whole image corresponds to domain  $[-1, 1]^2$ ; see also Figure SM1. Design  $A$  is constructed using our implementation of Siddon’s algorithm (original paper [SM28]) for parallel beam geometry with 64 projections and 64 parallel lines per projection.

**SM6. Assumptions on  $\mathcal{PP}$ . Nonparametric constructions of  $\pi_{\mathcal{M}}$  and  $\pi_{\mathcal{M}}(\cdot|Y^t \vee Z^t, t)$ .** Assumption in (4.10) is not very restrictive in view of the physical model of ET. According to Theorem 2.4.V, [SM9], any point process  $\mathcal{PP}$  (on a Polish space) which is a.s. boundedly-finite, has independent increments (events in different LORs and time periods are independent) and also orderly (no batches, i.e., no more than one point is possible a.s. in an infinitesimal set), is a Poisson point process. Therefore, at worst our assumption in (4.10) restricts  $\mathcal{PP}$  from the family of non-stationary Poisson processes to temporally stationary ones. In reality, of course it better to model  $\mathcal{PP}$  as non-stationary. For this, first, it would require to change our initial model and the log-likelihood in (4.4) to the non-stationary one (if model is set temporal-stationary it is not coherent to consider prior on non-stationary processes and vice-versa). Second, we would need to include complicated models of tracer kinetics inside the body which are outside of the scope of this work. Nevertheless, with a few slight changes in subsection 4.1 one may derive NPL for ET also for non-stationary scenario. Finally, being completely outside of the scope of ET, any point process with independent increments (seen as a generator of random discrete measures) can be constructed via a Poisson process due to the celebrated theorem of Kingman (1967) (see [SM5], Section 10.1, Theorem 10.1.III.). Thus, defining a prior on  $\mathcal{PP}$  is equivalent putting a prior on the parameter measure of a certain Poisson process.

Derivation for non-parametric Poisson and gamma processes in subsection 4.4 essentially does not change – prior on  $\Lambda$  is defined via the same MGP construction as in (4.11) where  $\Lambda|\Lambda_{\mathcal{M}} \sim G(\theta^t \Lambda_{\mathcal{M}}, (\theta^t)^{-1} \cdot \mathbb{1}_Z)$ ,  $\mathbb{1}_Z$  is the identity function on  $Z$ . Posterior in (4.13) is also an

<sup>1</sup>Source code in Python of the experiment on the Gibbs-sampler can be found at <https://gitlab.com/eric.barat/npl-pet>.

MGP due to the following result (our adaptation of Theorem 3.1 from [SM19]):

**Theorem SM6.1.** *Let  $Y^t \sim \mathcal{PP}_\Lambda^t$  and  $G_{\alpha,\beta}$  be the prior on  $\Lambda$ . Then, the posterior distribution of  $\Lambda$  is a weighted gamma process  $G_{\alpha+Y^t, \frac{\beta}{1+t\beta}}$ .*

*Proof.* Claim follows directly from the result of Theorem 3.1 from [SM19]. Indeed, having iid samples  $N_1, \dots, N_n$  from a Poisson point process with intensity  $\nu$  is equivalent having sample  $N_1 + \dots + N_n$  for intensity  $n\nu$ . Therefore, parameter  $n$  is a direct analog of  $t$  in our considerations. Moreover, it is trivial to check that all results from Section 3 of [SM19] hold for  $n$  being replaced with  $t$ .

Theorem is proved. ■

**SM7. GEM-type algorithm.** Recall that the attractiveness of Algorithm 4 relies on having an efficient procedure for minimizing  $L_p(\lambda \mid \tilde{\Lambda}_b^t, A, 1, \beta^t/t)$  and  $L(\lambda_{\mathcal{M}} \mid \tilde{\Lambda}^t, A_{\mathcal{M}}, 1)$ . For integer-valued data  $Y^t \in \mathbb{N}_0^d$  the  $L_p(\lambda \mid Y^t, t)$  coincides with the penalized negative log-likelihood for Poisson-type sample and in this situation, provided penalty  $\varphi(\lambda)$  satisfies elementary conditions (convex,  $C^2$  – smooth), fast monotonic GEM-type algorithms [SM8], [SM32] can be used. In our setting intensities  $\tilde{\Lambda}^t, \tilde{\Lambda}_b^t$  are not integer-valued anymore, hence the GEM derivation machinery must be re-verified. We claim that the same so-called “GEM-type” iterative algorithms can be derived outside the context of a Poisson model and missing data. First, notice that EM belongs to the class of optimization transfer algorithms [SM17] also denoted as MM (Majoration Minimization) – the  $E$ -step is interpreted as the construction of a majorizing surrogate for the objective function,  $M$ -step corresponds to its consequent minimization. Using the convexity argument from [SM6] we construct the same majorizing surrogate for  $L(\lambda \mid \tilde{\Lambda}_b^t, A, 1)$  as in [SM8] in a completely algebraic way but now for arbitrary nonnegative term  $\tilde{\Lambda}_b^t$ . Further extension to  $L_p(\lambda \mid \tilde{\Lambda}_b^t, A, 1, \beta^t/t)$  is straightforward by considering a separate surrogate for  $\varphi(\lambda)$ . An immediate and substantial consequence for practitioners is that all celebrated GEM algorithms for MLE and MAP reconstructions can be used in the bootstrap context by simply replacing data term by  $\tilde{\Lambda}_b^t$ .

**SM7.1. Majorizing surrogate of  $L(\lambda \mid \tilde{\Lambda}_b^t, A, 1)$ .** In [SM6] authors propose a purely algebraic derivation of the surrogate which we adapt below.

Let  $f_i(x) \triangleq x - \tilde{\Lambda}_{b,i}^t \log(x)$ ,  $\lambda_j^{(r)} \succeq 0$ ,  $j = 1, \dots, p$ , be the  $r^{\text{th}}$  iterate of the optimization algorithm minimizing  $L(\lambda \mid \tilde{\Lambda}_b^t, A, 1)$ , and denote also  $\Lambda_i^{(r)} = a_i^T \lambda^{(r)}$ . Consider the formula

$$\begin{aligned} L(\lambda \mid \tilde{\Lambda}_b^t, A, 1) &= \sum_{i=1}^d f_i(\Lambda_i) \\ &= \sum_{i=1}^d f_i \left( \sum_{j=1}^p a_{ij} \lambda_j \right) \\ &= \sum_{i=1}^d f_i \left( \sum_{j=1}^p \left[ \frac{a_{ij} \lambda_j^{(r)}}{\Lambda_i^{(r)}} \right] \left[ \frac{\lambda_j}{\lambda_j^{(r)}} \Lambda_i^{(r)} \right] \right). \end{aligned}$$

Since  $f_i$  is convex for  $\tilde{\Lambda}_{b,i}^t \geq 0$  and  $\sum_{j=1}^p \frac{a_{ij}\lambda_j^{(r)}}{\Lambda_i^{(r)}} = 1$ , using the Jensen's inequality we obtain

$$L(\lambda \mid \tilde{\Lambda}_b^t, A, 1) \leq Q_L(\lambda, \lambda^{(r)}),$$

where

$$Q_L(\lambda, \lambda^{(r)}) \triangleq \sum_{i=1}^d \sum_{j=1}^p \left[ \frac{a_{ij}\lambda_j^{(r)}}{\Lambda_i^{(r)}} \right] f_i \left( \frac{\lambda_j}{\lambda_j^{(r)}} \Lambda_i^{(r)} \right).$$

Note also that  $Q_L(\lambda^{(r)}, \lambda^{(r)}) = L(\lambda^{(r)} \mid \tilde{\Lambda}_b^t, A, 1)$ , hence,  $Q_L(\lambda, \lambda^{(r)})$  is indeed a surrogate for  $L(\lambda \mid \dots)$ . Using the definition of  $f_i$  we find that

$$\begin{aligned} Q_L(\lambda, \lambda^{(r)}) &= \sum_{i=1}^d \sum_{j=1}^p \left[ a_{ij}\lambda_j - \frac{a_{ij}\lambda_j^{(r)}}{\Lambda_i^{(r)}} \tilde{\Lambda}_{b,i}^t \log \left( \frac{\lambda_j}{\lambda_j^{(r)}} \Lambda_i^{(r)} \right) \right] \\ &= \sum_{j=1}^p A_j \left[ \lambda_j - \left( \frac{\lambda_j^{(r)}}{A_j} \sum_{i=1}^d \frac{a_{ij}\tilde{\Lambda}_{b,i}^t}{\Lambda_i^{(r)}} \right) \log \lambda_j \right] + R_L^{(r)}, \end{aligned}$$

where  $R_L^{(r)}$  is the remainder independent of  $\lambda$ . Function  $Q_L(\lambda, \lambda^{(r)})$  can be rewritten as follows:

$$(SM7.1) \quad Q_L(\lambda, \lambda^{(r)}) = \sum_{j=1}^p A_j (\lambda_j - \lambda_j \log \lambda_j) + R_L^{(r)}.$$

Hence, the update of  $\lambda_L^{(r+1)}$  by minimizing  $Q_L(\lambda, \lambda^{(r)})$  takes the following celebrated form:

$$(SM7.2) \quad \lambda_{j,L}^{(r+1)} \triangleq \frac{\lambda_j^{(r)}}{A_j} \sum_{i=1}^d \frac{a_{ij}\tilde{\Lambda}_{b,i}^t}{\Lambda_i^{(r)}}.$$

**SM7.2. Majorizing surrogate for  $\varphi(\lambda)$ .** Let

$$\varphi(\lambda) = \sum_{j=1}^p \sum_{k \in \mathcal{N}_j} w_{jk} \psi(\lambda_j - \lambda_k),$$

where  $w_{jk} > 0$ ,  $w_{kj} = w_{jk}$  are the weights and  $\mathcal{N}_j$  denotes the neighborhood of pixel  $j$ .

From [SM7], if

- i.  $\psi$  is symmetric.
- ii.  $\psi$  is continuous and differentiable everywhere.
- iii.  $\psi$  is convex.
- iv.  $\omega_\psi(u) \triangleq \frac{1}{u} \frac{d\psi(u)}{du}$  is non-increasing for  $u \geq 0$ .
- v.  $\lim_{u \rightarrow 0} \omega_\psi(u)$  is finite and positive,

then,  $\varphi$  can be majorized by a parabolic curve. In our work the above conditions are satisfied, therefore  $\varphi(\lambda)$  is majorized by a separable quadratic penalty given below (see [SM32] and references therein):

$$\varphi(\lambda) \leq Q_\varphi(\lambda; \lambda^{(r)}),$$

where

$$(SM7.3) \quad Q_\varphi(\lambda; \lambda^{(r)}) = \frac{1}{2} \sum_{j=1}^p p_{j,\varphi}^{(r+1)} (\lambda_j - \lambda_j^{(r)})^2,$$

$$(SM7.4) \quad p_{j,\varphi}^{(r+1)} = 4 \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(\lambda_j^{(r)} - \lambda_k^{(r)}),$$

so the iterative update for minimization is given by the formula:

$$(SM7.5) \quad \lambda_{j,\varphi}^{(r+1)} = \frac{2}{p_{j,\varphi}^{(r+1)}} \sum_{k \in \mathcal{N}_j} w_{jk} \omega_\psi(\lambda_j^{(r)} - \lambda_k^{(r)}) (\lambda_j^{(r)} + \lambda_k^{(r)}).$$

**SM7.3. Global surrogate minimization.** At iteration  $(r + 1)$ , solving the Karush-Kuhn-Tucker condition for minimizing the combined surrogate, we get

$$\lambda^{(r+1)} = \arg \min_{\lambda \geq 0} Q_L(\lambda, \lambda^{(r)}) + \frac{\beta^t}{t} Q_\varphi(\lambda, \lambda^{(r)}),$$

which gives a unique analytical positive solution

$$(SM7.6) \quad \lambda_j^{(r+1)} = \frac{2\lambda_{j,L}^{(r+1)}}{\sqrt{(b_j^{(r+1)})^2 + 4\beta_j^{(r+1)}\lambda_{j,L}^{(r+1)}} + b_j^{(r+1)}},$$

where

$$\beta_j^{(r+1)} = \frac{\beta^t}{t A_j} p_{j,\varphi}^{(r+1)} \quad \text{and} \quad b_j^{(r+1)} = 1 - \beta_j^{(r+1)} \lambda_{j,\varphi}^{(r+1)}.$$

The GEM-type algorithm is summarized in [Algorithm SM1](#).

*Remark SM7.1.* By setting  $\frac{\beta^t}{t} = 0$  in (SM7.6), we find immediately that  $\lambda^{(r+1)} = \lambda_L^{(r+1)}$ .

*Remark SM7.2.* Parameter  $\tilde{\lambda}_{\mathcal{M}}^t$  in [Algorithm 3](#) is easily obtained by iterating formula (SM7.2) with projector  $A_{\mathcal{M}}$  and random intensities  $\tilde{\Lambda}^t$

$$(SM7.7) \quad \tilde{\lambda}_{\mathcal{M},s}^{t,(r+1)} = \frac{\tilde{\lambda}_{\mathcal{M},s}^{t,(r)}}{A_{\mathcal{M},s}} \sum_{i=1}^d \frac{a_{\mathcal{M},is} \tilde{\Lambda}_i^t}{\tilde{\Lambda}_{\mathcal{M},i}^{t,(r)}}.$$

## SM8. Details of the numerical experiment of [Algorithm 4](#). <sup>2</sup>

---

<sup>2</sup>Source code in Python of the experiment can be found at <https://gitlab.com/eric.barat/npl-pet>

---

**Algorithm SM1**  $\arg \min_{\lambda \succeq 0} L_p(\lambda \mid \tilde{\Lambda}_b^t, A, 1, \frac{\beta^t}{t})$  by optimization transfer

---

1: **data** : intensities  $\tilde{\Lambda}_b^t$ ;  
2: **input** : Initial image  $\lambda^{(0)}$ , number max. of iterations  $R$ , projector  $A$ , regularization parameter  $\beta^t$ , penalty  $\varphi(\lambda)$   
3: **for**  $r \leftarrow 0$  to  $R - 1$  **do**  
4:   **for**  $j \leftarrow 1$  to  $p$  **do**  
5:     compute  $\lambda_{j,L}^{(r+1)}$  using (SM7.2)  
6:     compute  $\lambda_{j,\varphi}^{(r+1)}$  using (SM7.5)  
7:     compute  $\lambda_j^{(r+1)}$  using (SM7.6)  
8:   **end for**  
9: **end for**  
10: **return**  $\lambda^{(R)}$

---

**SM8.1. Choice of penalty  $\varphi$ .** We take the well-known in PET imaging log cosh penalty [SM12] coupled with  $\ell_2$ -squared pairwise difference term:

$$(SM8.1) \quad \varphi(\lambda) = \sum_{j=1}^p \sum_{j' \in \mathcal{N}_j} w_{jj'} \left( (1 - \nu) \zeta \log \cosh \left( \frac{\lambda_j - \lambda_{j'}}{\zeta} \right) + \frac{\nu}{2} (\lambda_j - \lambda_{j'})^2 \right),$$

where  $w_{jj'} > 0$ ,  $w_{j,j} = w_{j,j'}$  and  $\mathcal{N}_j$  the neighborhood of pixel  $j$  consisting of 8-adjacent ones –  $w_{jj'} = 1$  for horizontal/vertical neighbors and  $w_{jj'} = \frac{\sqrt{2}}{2}$  for diagonal ones.

Penalty in (SM8.1) is attractive since it bridges together Gaussian prior for pairwise interactions (when  $\zeta \rightarrow +\infty$ ), and for  $\nu = 0$ ,  $\zeta \rightarrow 0$ , it corresponds to pairwise  $\ell^1$ -penalty (Laplace prior). It is easy to check that  $\varphi(\lambda)$  in (SM8.1) is strictly convex except the only direction spanned on vector  $e = (1, \dots, 1)$ . From (2.5) it follows that  $e \notin \ker A$ , therefore conditions (2.11) and (2.12) are automatically satisfied. Recall that for our experiments we choose  $\zeta$ ,  $\nu$  be always fixed and equal  $\zeta = 0.05$ ,  $\nu = 0.15$ .

**SM8.2. Calibration assessment and metrics.** Following [SM30], we say that a model is perfectly calibrated if for all levels  $\alpha \in [0, 1]$ , the corresponding marginal pixel-wise distributions contain  $\lambda_{*opt,j}$ 's in credible intervals with confidence level  $\alpha$  for  $\alpha \cdot 100\%$  of pixels. For each pixel  $1 \leq j \leq p$ , the credible interval  $\mathcal{I}_{j,\alpha} = (q_{j,\alpha}^L, q_{j,\alpha}^U]$  of confidence level  $\alpha$  is defined as:

$$(SM8.2) \quad P(\tilde{\lambda}_{b,j}^t \in \mathcal{I}_{j,\alpha} \mid Y^t) = \alpha$$

Thus, the perfect calibration means that (modulo numerical discretization effects):

$$(SM8.3) \quad \frac{1}{p} \sum_{j=1}^p \mathbb{1}\{\lambda_{*opt,j} \in \mathcal{I}_{j,\alpha}\} = \alpha, \quad \forall \alpha \in [0, 1].$$

As in [SM15], we choose  $\mathcal{I}_{j,\alpha}$  to be the smallest credible interval in terms of Lebesgue measure on  $\mathbb{R}$  which is also known as *highest posterior density* (HPD) region. Such choice



(e.g., rather than equal-tailed interval) is especially well-motivated in the ET context since the posterior for  $\tilde{\lambda}_{b,j}^t \in \mathbb{R}_+$  could be asymmetric for intensities near the zero-boundary. For any  $j$ , we compute empirical HPD-intervals from  $B$  samples of NPL, denoted by  $\hat{\mathcal{I}}_{j,\alpha} = (\hat{q}_{j,\alpha}^L, \hat{q}_{j,\alpha}^U]$ .

Since both  $p$  and  $B$  are finite, confidence levels for HPD-intervals must be discretized, so the computed (or empirical) *achieved* expected coverage (for  $\lambda_{*opt}$ ) is given by the formula:

$$(SM8.4) \quad C_A\left(\frac{m}{M}; \lambda_{*opt}\right) \triangleq \frac{1}{p} \sum_{j=1}^p \mathbb{1}\{\lambda_{*opt,j} \in \hat{\mathcal{I}}_{j,\frac{m}{M}}\}, \quad m \in \{1, \dots, M\},$$

where  $M$  is the size of the uniform grid on  $[0, 1]$ . The *targeted* expected coverage coincides with the confidence level for perfect calibration and equals  $\frac{m}{M}$  for each  $m \in \{1, \dots, M\}$ .

Note that  $\hat{\mathcal{I}}_{j,\alpha}$  is not random (being conditioned on  $Y^t$ ) and depends on  $\rho$ . Therefore, the goal of calibration is to tune  $\rho$  in order to satisfy (SM8.3) (modulo discretizations) as close as possible. To quantify the error we use the *expected calibration error* (ECE) metric which is defined as follows (we omit the dependence on  $\rho$  for simplicity of notations):

$$\text{ECE} \triangleq \frac{1}{M} \sum_{i=1}^M \left| C_A\left(\frac{m}{M}; \lambda_{*opt}\right) - \frac{m}{M} \right|.$$

Interestingly,  $C_A(\cdot; \lambda_{*opt})$  can be interpreted as an empirical c.d.f since

$$(SM8.5) \quad C_A(x; \lambda_{*opt}) = \frac{1}{p} \sum_{j=1}^p \mathbb{1}\{l_j \leq x\}, \quad l_j(\lambda_{*opt}) = \min_{\alpha \in [0,1]} \{\alpha \mid \lambda_{*opt} \in \hat{\mathcal{I}}_{j,\alpha}\}.$$

Then, in fact, ECE in (SM8.4) coincides with the 1-Wasserstein distance between uniform distribution and the empirical distribution of  $\{l_j\}_{j=1}^p$ . Values  $\{l_j\}_{j=1}^p$  can be seen as a result of evaluation of NPL-HPD-forecasters at  $\{\lambda_{*opt,j}\}_{j=1}^p$ , respectively (since the returned probability is not a value of the conditional c.d.f but a minimal level of the HPD-region containing  $\lambda_{*opt,j}$ ); see [?] and its Supplementary Materials, Proposition 3. The spatial representation of  $\{l_j\}_{j=1}^p$  constitutes what we call the coverage map computed by the formula:

$$\hat{l}_j = \min_{m \in \{1, \dots, M\}} \left\{ \frac{m}{M} \mid \lambda_{*opt,j} \in \hat{\mathcal{I}}_{j,\frac{m}{M}} \right\}, \quad j \in \{1, \dots, p\}.$$

Because for the perfect calibration  $C_A(\cdot; \lambda_{*opt})$  should be close to c.d.f of the uniform distribution we consider its empirical p.d.f. and call it coverage histogram in Figure 6(e) (in view of the above representation in (SM8.5) via  $\{l_j\}_{j=1}^p$ ). We quantify the overall discrepancy between coverage histogram and uniform density by computing the KL-divergence:

$$\text{KLC} \triangleq -\frac{1}{M} \sum_{m=1}^m \log(M \cdot h_m),$$

where  $h_m$  is the  $m^{\text{th}}$ -bin value of the normalized histogram of  $\hat{l}_j$  ( $\sum_{m=1}^M h_m = 1$ ).

We finally consider two additional metrics for assessing the predictive accuracy – PSNR and MSWD (*peak-signal-to-noise-ratio* and the *mean-squared-weighted-deviation*):

$$\text{PSNR}(\lambda_{*opt}, \bar{\lambda}^t) = 10 \log_{10} \left( \frac{p \cdot \max(\lambda_{*opt})^2}{\sum_{j=1}^p (\lambda_{*opt,j} - \bar{\lambda}_j^t)^2} \right), \bar{\lambda}^t = \frac{1}{B} \sum_{b=1}^B \tilde{\lambda}_b^t,$$

$$\text{MSWD}(\lambda_{*opt}, \bar{\lambda}_b^t) = \frac{1}{p} \sum_{j=1}^p \frac{(\lambda_{*opt,j} - \bar{\lambda}_j^t)^2}{\sigma_j^2}, \sigma_j^2 = \frac{1}{B} \sum_{b=1}^B (\tilde{\lambda}_{b,j}^t - \bar{\lambda}_j^t)^2.$$

Metric MSWD has form of the chi-squared statistic under the assumption that indices  $j$  denoting different pixels correspond to independent observations with normal distributions. As a rule of thumb one says that for  $\text{MSWD} \approx 1.0$  the analytical model estimates optimally (in the sense of the above assumption) the uncertainties in each pixel, and for  $\text{MSWD} < 1.0$  and  $\text{MSWD} > 1.0$  posterior overestimates and underestimates the latter, respectively.

*Remark SM8.1.* Note that by performing our assessment on one single  $\lambda_{*opt}$ , we implicitly assume that  $\lambda_{*opt}$  is sufficiently representative for a large family of isotope concentrations in the patient. Certainly, for a more profound numerical investigation one should try to calibrate against large family of realistic  $\lambda_{*opt}$  simultaneously. Another important observed feature (specific for ET) is the effect of huge amount of nearly null-intensity pixels outside of the cranium (nearly null for  $\lambda_{*opt}$  and exactly null for  $\lambda_*$ ), which can significantly bias the coverage analysis. Indeed, many LORs crossing such pixels we have  $Y_i^t \equiv 0$  a.s. and contraction of the NPL posterior to a degenerate  $\delta_0$ -distribution is very fast making the overall model extra-conservative for low levels of the expected coverage; see the calibration curve in [Figure SM2](#). As a consequence, these nearly zero values (at machine precision) of  $\lambda_{*opt,j}$  are captured by any small confidence level HPD. To mitigate this bias, we restrict the contributing pixels to ones inside the discrete convex-hull defined by  $\lambda_{*,j} > 0$  which is very close to pixels inside the cranium boundary; see the grey mask, for example, in [Figure 6\(d\)](#). Note also that applying such mask (to remove exterior of the patient’s body from the model) is a common technique in ET since the latter can be directly computed from MRI or CT scans and, in principle, does not require  $\lambda_*$ .

**SM8.3. Additional results for large  $t$ .** To test the hypothesis on overcontraction in the slab between the cranium and the soft brain tissue we increased  $t$  up to  $t = 100$ . Thus, the LORs crossing this region which previously received  $Y_i^t = 0$  for  $t = 1$  but having  $\Lambda_i^* > 0$  will have much greater chances to receive  $Y_i^t > 0$ . To mitigate possible bias due to regularizer  $\varphi$  we set  $\beta^t/t = \beta = \beta_{min} = 1 \cdot 10^{-3}$  (same as for computing  $\lambda_{*opt}$ ). In [Figure SM3](#) one can see that the model is almost perfectly calibrated and the aforementioned slab is no longer contrasted as it was in [Figure 6\(d\)](#) which is predicted by our initial explanation. Slight overconfidence is visible on the calibration curve (c) and on the coverage histogram (e) for small confidence levels for which we yet have no explanation apart of numerical errors of machine precision/low number of samples/small value of  $t$ .

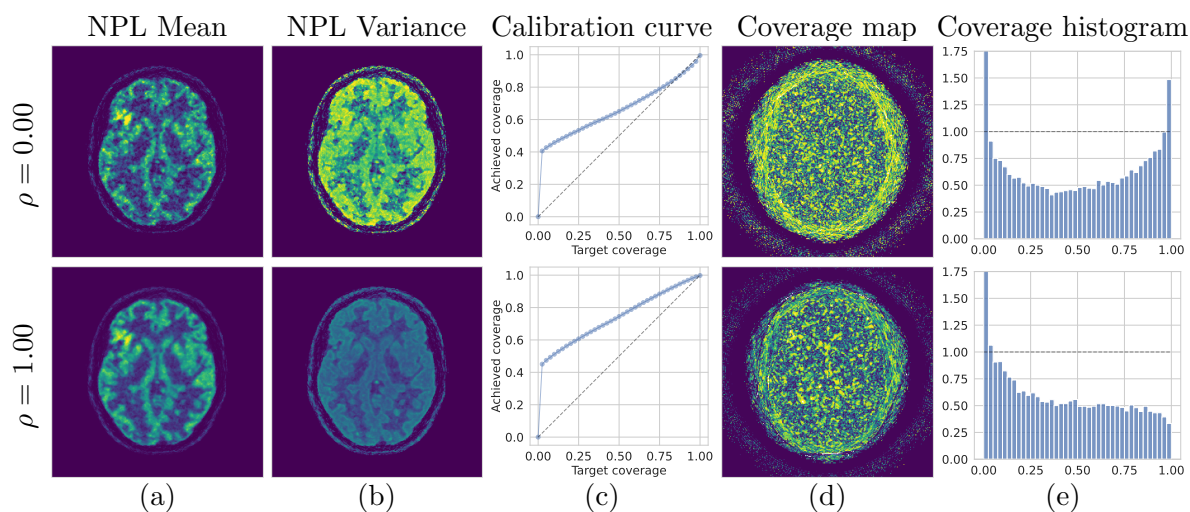
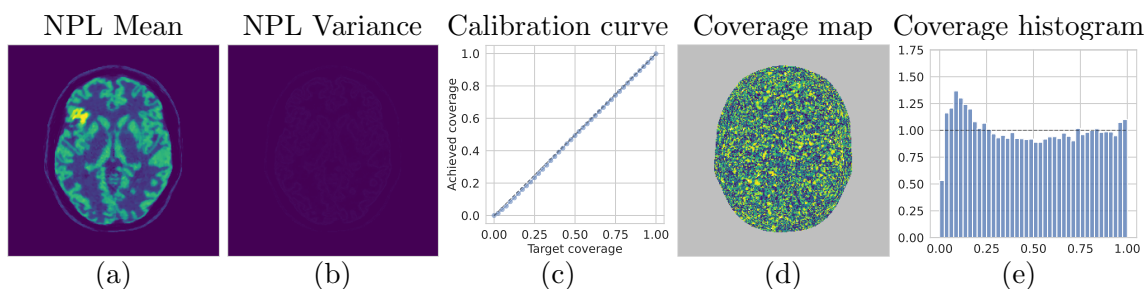


Figure SM2: NPL calibration without mask for pixels outside the cranium


 Figure SM3: NPL with  $\rho = 0$  and  $\beta = \beta_{min}$  for  $t = 100$ 

We also additionally checked that NPL posterior mean for  $\rho = 0$  (no MRI) is indistinguishable from the MAP reconstruction with the same penalty tuning; see [Figure SM4](#). This is specifically attractive for practitioners in ET who are acquainted with MAP estimates but much less with posterior mean from some complicated posteriors as in NPL. The result also supports the claim in [Theorem 6.10](#) that asymptotically posterior concentrates primarily not around  $\lambda_{*opt}$  but a strongly consistent estimator  $\hat{\lambda}_{sc}^t$  for which we conjectured to coincide asymptotically, for example, with the MAP estimate; see [subsection 6.3](#).

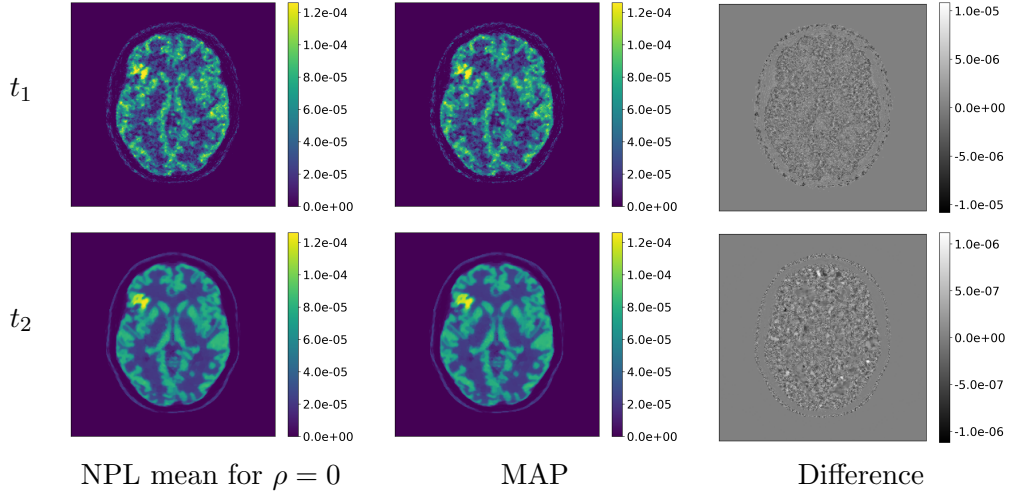


Figure SM4: NPL mean for  $\rho = 0$  compared to MAP reconstruction for  $t_1 = 1$  and  $t_2 = 100$ ;  $\beta^t = 2 \times 10^{-3}$ ;  $B = 8192$

From the two simulations for  $t_1 = 1$  and  $t_2 = 100$  one may observe that the empirical contraction rate of absolute differences seems to be of order  $t^{-1/2}$ . This can be explained by the fact that for regular models with  $n$  i.i.d observations (recall that model of ET is regular for pixels where  $\lambda_{*,j} > 0$ ), the next error term beyond the normal approximation in the first order Edgeworth's expansion of the posterior decays with rate  $n^{-1/2}$ , which is  $t^{-1/2}$  in our parametrization; see [SM25].

**SM9. Discussion on bayesian approach and NPL in ET.** By our knowledge the only yet existing theoretically-oriented work on the bayesian model of ET is [SM4]. The non-regular model considered there is exactly as in (2.1), being also well-specified and with injective design  $A$  (though ill-conditioned). An essential common thing is that both posteriors (bayesian and NPL) experience splitting into three modes due to positivity constraints. In [SM4] asymptotic bayesian posterior has exponential distribution for pixels which are intersected by at least one ray s.t.  $\Lambda_i^* = 0$  ( $p(\lambda_j|Y^t, t) \approx \text{Exp}(-tA_{*j})$ ,  $A_{*j} = \sum_{i:\Lambda_i^*=0} a_{ij}$ ; union of indices for such pixels being denoted by  $S_1$ ) and truncated gaussian distributions to  $\mathbb{R}_+^p$  for the rest (denoted by  $S_0 = \overline{S_1}$ ):

$$(SM9.1) \quad \text{approximately for large } t: \sqrt{t}(\lambda_{S_0} - \lambda_{*,S_0})|Y^t, t \sim \mathcal{TN}(a_0, \Omega_{00}^{-1}),$$

$$(SM9.2) \quad \Omega_{00} = A_{I_1(\Lambda^*), S_0}^T \text{diag}(1/\Lambda_i^*; i \in I_1(\Lambda^*)) A_{I_1(\Lambda^*), S_0},$$

$$(SM9.3) \quad a_0 = \Omega_{00}^{-1} A_{I_1(\Lambda^*), S_0}^T \left[ \frac{Y_i - t\Lambda_i^*}{\sqrt{t\Lambda_i^*}} \right]_{I_1(\Lambda^*)},$$

where  $\mathcal{TN}$  denotes the truncated gaussian distribution such that  $\lambda_{S_0}|Y^t, t \in \mathbb{R}_+^p$  (it is pure gaussian distribution where positive constraints are a.s. inactive (also regular part of the model) and half-gaussian where the latter are asymptotically active with high probability); see Section 6 for details.

For NPL, yet we do not prove asymptotic normality result but only tightness in [Theorem 6.10](#), however, a very similar splitting is visible. Indeed, linear space  $\mathcal{V}$ , defined in [\(6.16\)](#), coincides with span of pixels in  $S_1$  in [\[SM4\]](#) and we prove that the contraction rate of the NPL-posterior is faster in directions of  $\mathcal{V}$  than in others (see formulas [\(6.24\)](#) and [\(6.25\)](#)) same as the exponential part for the bayesian posterior above has faster contraction rate than the truncated gaussian part in [\(SM9.1\)](#)-[\(SM9.3\)](#). For simplicity, if in our model we assume that  $A$  is injective, than  $\mathcal{W} = \{0\}$  (see formula [\(6.18\)](#)) and  $\Pi_{\mathcal{U}}\lambda = \lambda_{S_0}$ , where  $\mathcal{U}$ ,  $\Pi_{\mathcal{U}}$  are defined in [\(6.17\)](#), [\(6.19\)](#), respectively. Thus, our result in [\(6.25\)](#) can be seen as a preliminary step towards the demonstration of asymptotic normality along  $\mathcal{U}$  whose bayesian counterpart is given in [\(SM9.1\)](#)-[\(SM9.3\)](#).

Now we discuss [Conjecture 6.11](#) and the intuition why it should hold, especially, in view of results from [\[SM4\]](#). First, in [\(SM9.3\)](#) not that and

$$(SM9.4) \quad \Lambda^* = A_{I_1(\Lambda^*), S_0} \lambda_{*, S_0}.$$

Then, using the above formula and [\(SM9.1\)](#) we get:

$$(SM9.5) \quad \text{approximately for large } t: \sqrt{t}\lambda_{S_0}|Y^t, t \sim \mathcal{TN}(\xi, \Omega_{00}^{-1}),$$

$$(SM9.6) \quad \xi = \sqrt{t} \cdot \Omega_{00}^{-1} A_{I_1(\Lambda^*), S_0}^T \frac{Y_i/t}{\Lambda_i^*}.$$

Formulas [\(SM9.5\)](#), [\(SM9.6\)](#) and the fact that asymptotic posterior in direction  $\mathcal{V}$  has exponential distribution imply that the bayesian MAP estimate from [\[SM4\]](#) is a perfect candidate for  $\hat{\lambda}_{sc}^t$  in [Conjecture 6.11](#). Indeed, for such choice  $\Pi_{\mathcal{V}}\hat{\lambda}_{sc}^t \approx 0$ ,  $\Pi_{\mathcal{U}}\hat{\lambda}_{sc}^t \approx \xi/\sqrt{t}$  (modulo truncation for positivity), hence, conditions [\(6.22\)](#), [\(6.23\)](#) are automatically satisfied. To be rigorous, one must carefully check that the above approximations are valid up to terms decaying with faster rates.

### SM10. Remark on centering term of the posterior.

**Definition SM10.1.** *We say that  $U^t$  converges in conditional distribution to  $V$  almost surely  $Y^t$ ,  $t \in [0, +\infty)$  if for every Borel set  $A \in B(\mathbb{R}^n)$  the following holds:*

$$(SM10.1) \quad P(U^t \in A \mid Y^t) \rightarrow P(V \in A) \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty).$$

*This type of convergence will be denoted as follows:*

$$(SM10.2) \quad U^t \xrightarrow{c.d.} U.$$

Centering the distribution of  $\tilde{\lambda}_b^t$  at the true parameter  $\lambda_*$  in (ii) does not allow to achieve conditional tightness almost surely  $Y^t$ ,  $t \in (0, +\infty)$  which we briefly explain below.

As a part of the proof of [Theorem 6.10](#) (see [Lemmas SM12.11](#) and [SM12.12](#)) we show that

$$(SM10.3) \quad \Pi_{\mathcal{U}}(\tilde{\lambda}_b^t - \hat{\lambda}_{sc}^t) - u^t(\tilde{\xi}^t) \xrightarrow{c.p.} 0,$$

where

$$(SM10.4) \quad \tilde{\xi}^t = (\dots, \sqrt{t} \frac{\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t}{\sqrt{\hat{\Lambda}_{sc,i}^t}}, \dots), \quad i \in I_1(\Lambda^*),$$

$$(SM10.5) \quad u^t(\xi) = \arg \min_{\substack{u: (1-\Pi_{\mathcal{V}})\hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + w \succeq 0, \\ u \in \mathcal{U}, w \in \mathcal{W}}} -u^T (A_{I_1(\Lambda^*)})^T (\hat{D}_{I_1(\Lambda^*)}^t)^{-1/2} \xi + \frac{1}{2} u^T \hat{F}_{I_1(\Lambda^*)}^t u,$$

$$(SM10.6) \quad \hat{D}_{I_1(\Lambda^*)}^t = \text{diag}(\dots, \hat{\Lambda}_{sc,i}^t, \dots), \quad i \in I_1(\Lambda^*),$$

$$(SM10.7) \quad \hat{F}_{I_1(\Lambda^*)}^t = \sum_{i \in I_1(\Lambda^*)} \frac{a_i a_i^T}{\hat{\Lambda}_{sc,i}^t} = (A_{I_1(\Lambda^*)})^T (\hat{D}_{I_1(\Lambda^*)}^t)^{-1} A_{I_1(\Lambda^*)}.$$

That is the conditional tightness (and also the asymptotic distribution) of  $\Pi_{\mathcal{U}}(\tilde{\lambda}_b^t - \hat{\lambda}_{sc}^t)$  asymptotically coincides with  $u^t(\tilde{\xi}^t)$  being the minimizer of a quadratic function on a polyhedral set depending on  $\hat{\lambda}_{sc}^t$ . In the proof we show that conditional tightness of  $u^t(\tilde{\xi}^t)$  is implied by tightness of  $\tilde{\xi}^t$  (this is especially obvious if the constraints in (SM10.5) are not active for large  $t$ , e.g., when  $\lambda_* \succ 0$ ) and that under the assumptions of the theorem it holds that

$$(SM10.8) \quad (\dots, \sqrt{t} \frac{\tilde{\Lambda}_{b,i}^t - \frac{Y_i^t}{t}}{\sqrt{\hat{\Lambda}_{sc,i}^t}}, \dots) \xrightarrow{c.d.} \mathcal{N}(0, I),$$

$I$  – identity matrix of size  $\#I_1(\Lambda^*) \times \#I_1(\Lambda^*)$ .

From (SM10.5)–(SM10.8) and the Prohorov’s theorem on tightness of weakly convergent sequences or r.v.s, the asymptotic behavior (tightness, distribution) of  $u^t(\tilde{\xi}^t)$  is essentially depends on the term

$(\dots, \sqrt{t} \frac{\hat{\Lambda}_{sc,i}^t - \frac{Y_i^t}{t}}{\sqrt{\hat{\Lambda}_{sc,i}^t}}, \dots)$ ,  $i \in I_1(\Lambda^*)$ . For tightness this term needs to be asymptotically bounded for almost any trajectory  $Y^t$ ,  $t \in (0, +\infty)$ , which is exactly asked in (6.22) (in a slightly weakened form).

Now, if we center  $\tilde{\lambda}_b^t$  on  $\lambda_*$  one finds that  $\hat{\lambda}_{sc}^t$  must be replaced everywhere with  $\lambda_*$  in (SM10.5)–(SM10.8) and, most importantly, the latter term becomes  $(\dots, \frac{Y_i^t - t\Lambda_i^*}{\sqrt{t\Lambda_i^*}}, \dots)$  being asymptotically standard normal w.r.t  $Y^t \sim PP_{A,\lambda^*}^t$  (see section SM2). Therefore, the mean of the asymptotic distribution of  $\sqrt{t}\Pi_{\mathcal{U}}(\tilde{\lambda}_b^t - \lambda_*)$  depends on the trajectory of  $(Y_i^t - t\Lambda_i^*)/\sqrt{t\Lambda_i^*}$ ,  $i \in I_1(\Lambda^*)$ , which is almost surely unbounded infinitely often on  $t \in [0, +\infty)$  in view of the Law of Iterated Logarithm for  $Y^t$  (see formula (SM2.4)). Thus, the tightness for  $\sqrt{t}\Pi_{\mathcal{U}}(\tilde{\lambda}_b^t - \lambda_*)$  almost surely for any trajectory  $Y^t$ ,  $t \in [0, +\infty)$  is impossible. A very similar behavior for centering of the posterior distribution for weighted bootstrap was also observed in Theorem 3.3 from [SM23].

**SM11. MRI data, mask and the non-expansiveness condition.** Below we consider a geometrical interpretation of the non-expansiveness condition based on representation of designs

$A, A_{\mathcal{M}}$  as weighted Radon transforms over the space of discrete images (which are commonly used in practice). We show that failure of this condition implies presence of a segment in  $M \in \mathcal{M}$  which is badly aligned with respect to the convex hull of the tracer support. To avoid such situations in practice, we propose to preprocess MRI images before using them in the context of ET which is explained in the end of this section.

For simplicity, let  $k = 1$ , i.e.,  $\mathcal{M} = \{M\}$ , and let

$$(SM11.1) \quad \Gamma = \{\gamma_i\}_{i=1}^d \text{ be the set of rays available in the acquisition geometry.}$$

Assume that  $A = (a_{ij})$  is a discretized version of some weighted Radon transform on set of rays  $\Gamma$  with positive weight  $W$ , that is

$$(SM11.2) \quad a_{ij} = \int_{\gamma_i} W(x, \gamma_i) \mathbf{1}_j(x) dx, \quad i \in \{1, \dots, d\}, j \in \{1, \dots, p\},$$

$$(SM11.3) \quad W = W(x, \gamma), (x, \gamma) \in \mathbb{R}^2 \times TS^1, 0 < c \leq W \leq C,$$

where  $dx$  denotes the standard Lebesgue measure on ray  $\gamma_i$ ,  $\mathbf{1}_j(x)$  is the indicator function of pixel  $j$  on the image. Weight  $W(x, \gamma)$  is some known sufficiently regular function of spatial coordinates and oriented rays in  $\mathbb{R}^2$  which are parameterized by  $TS^1$  (tangent bundle of the unit sphere, see e.g., [SM21]). Projectors defined by the formulas of type (SM11.2), (SM11.3) are common in CT and ET practice; see e.g., [SM28], [SM14] (for example, in PET and SPECT weight  $W$  is used to model attenuation and nonuniform sensitivity of detectors; see e.g., [SM26], [SM24], [SM11]).

From (2.15) and (SM11.2) it follows that

$$(SM11.4) \quad A_{\mathcal{M}} = (a_{M, is}), a_{M, is} = \int_{\gamma_i} W(x, \gamma_i) \mathbf{1}_{M, s}(x) dx, \quad s \in S(M),$$

where  $\mathbf{1}_{M, s}(x)$  is the indicator function of segment  $s$  in image  $M$ .

Recall that  $\lambda_* \in \mathbb{R}_+^p$  is the discretized version of the real spatial distribution of the tracer and assume that  $\lambda_* \in \mathbb{R}_+^p$  is pixel-wise connected (i.e., between two arbitrary pixels with positive tracer uptake there is a path of pixels preserving the positivity of the signal; two pixels are considered neighbors if they share a full edge (see Figure SM5(a))). The above assumption is natural, for example, in the context of brain imaging when the tracer is distributed in the whole volume inside the cranium and only relative spatial variations are of practical interest.

**Definition SM11.1.** *Let  $\Gamma$  be the finite family of oriented rays in  $\mathbb{R}^2$ ,  $A$  be the projector defined by formulas (SM11.2) and (SM11.3),  $\lambda_* \in \mathbb{R}_+^p$ ,  $\lambda_* \neq 0$  and  $\lambda_*$  is pixel-wise connected. Consider  $\gamma_i \in \Gamma$  and assume that  $i \in I_0(A\lambda_*)$ . Then, support of  $\lambda_*$  lies completely in one of the closed half-spaces in  $\mathbb{R}^2$  separated from each other with ray  $\gamma_i$ . Let  $H(\lambda_*, \gamma_i)$  be such a closed half-space. Consider the discrete version of  $H(\lambda_*, \gamma_i)$  defined by the formula*

$$(SM11.5) \quad \text{DH}(\lambda_*; \gamma_i) = \{j \in \{1, \dots, p\} \mid \text{intersection between pixel } j \text{ and } H(\lambda_*, \gamma_i) \\ \text{is of non-zero Lebesgue measure on } \mathbb{R}^2\}.$$

Consider

$$(SM11.6) \quad \text{D}\mathring{H}(\lambda_*; \gamma_i) = \{j \in \text{DH}(\lambda_*, \gamma_i) \mid \text{intersection between pixel } j \text{ and ray } \gamma_i \\ \text{is of length zero}\}.$$

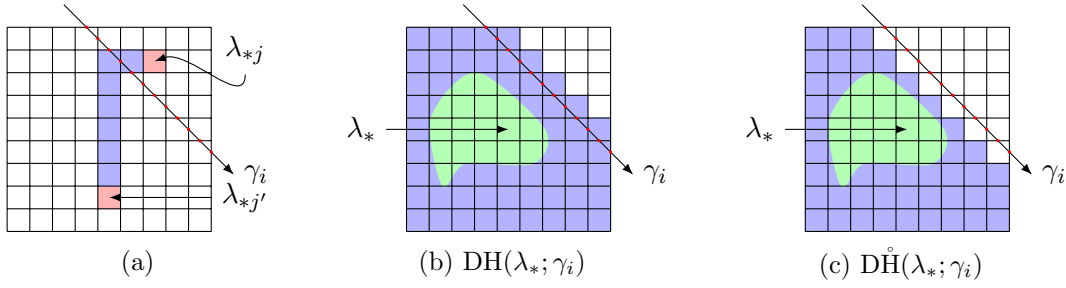


Figure SM5

Discrete convex hull of  $\lambda_*$  for family  $\Gamma$  is defined by the formula

$$(SM11.7) \quad DConv(\lambda_*; \Gamma, A\lambda_*) = \bigcap_{\substack{\gamma_i \in \Gamma, \\ i \in I_0(\Lambda_*)}} D\overset{\circ}{H}(\lambda_*; \gamma_i).$$

For the geometrical intuition behind definitions  $DH(\cdot)$ ,  $D\overset{\circ}{H}(\cdot)$ ,  $DConv(\cdot)$ , see examples (b), (c) in [Figure SM5](#).

Now assume that the non-expansiveness condition fails in the following sense:

$$(SM11.8) \quad \exists i \in I_0(\Lambda^*) \text{ such that } \Lambda_{\mathcal{M},i}^* > 0,$$

where  $\Lambda_{\mathcal{M}}^*$  is defined in (6.10). From (SM11.1)–(SM11.4) and [Definition SM11.1](#) it follows that in the image for  $\lambda_{\mathcal{M},*}$  there is a segment  $s \in S(M)$  which intersected by  $\gamma_i \in \Gamma$  and such that  $\lambda_{\mathcal{M},*,s} > 0$  (see [Figure SM6\(a\)](#)), that is

$$(SM11.9) \quad \bigcup_{\substack{M \in \mathcal{M}, \\ s \in S(M), \\ \lambda_{\mathcal{M},*,s} > 0}} s \not\subset DConv(\lambda_*; \Gamma, \Lambda^*).$$

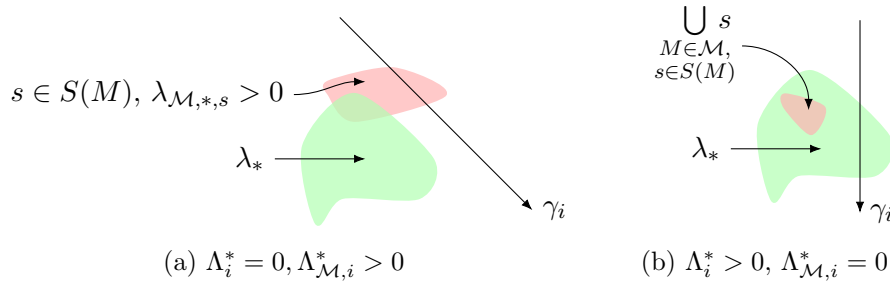


Figure SM6



If we assume that  $\lambda_{\mathcal{M},*}$  is also pixel-wise connected, then from (SM11.9) it follows that

$$(SM11.10) \quad \text{DConv}(\lambda_{\mathcal{M},*}; \Gamma, A_{\mathcal{M}}\lambda_{\mathcal{M},*}) \not\subseteq \text{DConv}(\lambda_*; \Gamma, A\lambda_*).$$

To conclude, we have just demonstrated the following statement.

**Proposition SM11.2.** *Let  $\lambda_* \in \mathbb{R}_+^p$ ,  $\lambda_* \neq 0$ ,  $\lambda_*$  is pixel-wise connected and designs  $A$ ,  $A_{\mathcal{M}}$  be of type (SM11.1)–(SM11.4). Let  $\lambda_{\mathcal{M},*}$  be a solution of the minimization problem in (6.9) and  $\lambda_{\mathcal{M},*}$  be also pixel-wise connected. Assume that the non-expansiveness condition (Assumption 6.7) fails in the sense of (SM11.8). Then, formula (SM11.10) holds.*

To avoid the situation in Proposition SM11.2 one may propose to use a significantly smaller segmentation area, for example, such that

$$(SM11.11) \quad \bigcup_{\substack{M \in \mathcal{M}, \\ s \in S(M)}} s \subsetneq \text{DConv}(\lambda_*; \Gamma, \Lambda^*),$$

where  $A \subsetneq B$  denotes the strict inclusion of sets. In this case even a small misalignment may lead to a situation when  $\mathcal{KL}(PP_{A,\lambda_*}^t, PP_{A_{\mathcal{M}},\lambda_{\mathcal{M}}}^t) = +\infty$ , so the KL-projection of  $PP_{A,\lambda_*}^t$  onto MRI-based model  $PP_{A_{\mathcal{M}},\lambda_{\mathcal{M}}}^t$  is impossible; see Figure SM6(b). In view of the latter an ideal choice for  $S(M)$  would be such that

$$(SM11.12) \quad \text{DConv}(\lambda_{\mathcal{M},*}; \Gamma, A_{\mathcal{M}}\lambda_{\mathcal{M},*}) = \text{DConv}(\lambda_*; \Gamma, A\lambda_*).$$

The above arguments can be easily extended to the case of  $k > 1$  by simply checking the alignments for all images in  $\mathcal{M}$  simultaneously. We conclude with a proposition to use the following pipeline for preprocessing of  $\mathcal{M}$ :

1. Estimate  $\text{DConv}(\lambda_*; \Gamma, A\lambda^*)$  using any well-suited and fast algorithm (and/or medical expertise). Let  $D$  be such an estimate.
2. In all MRI-images remove pixels lying outside of  $D$  and perform segmentations only on those which are left inside of  $D$ .

In view of steps 1, 2 above we propose an alternative name for Assumption 6.7 – mask condition. The term “mask” is also used in practice to denote restrictions of support of the tracer (e.g., using the medical expertise), so the above procedure theoretically reflects existing empirical practices.

## SM12. Proofs.

**Lemma SM12.1.** *Let  $\varphi(\lambda)$  be the function satisfying (2.11), (2.12),  $A$  satisfies conditions in (2.3)–(2.5). Let  $\lambda \in \mathbb{R}_+^p$  and  $U \subset \text{Span}(A^T)$  be a compact such that*

$$(SM12.1) \quad \{w : \lambda + u + w \succeq 0, w \in \ker A\} \text{ is non-empty for any } u \in U.$$

Then, mapping defined by the formula

$$(SM12.2) \quad w_{A,\lambda}(u) = \arg \min_{\substack{w: \lambda + u + w \succeq 0, \\ w \in \ker A}} \varphi(\lambda + u + w), \quad u \in U$$

is one-to-one. Moreover,  $w_{A,\lambda}(u)$  is continuous on  $U$ .

**SM12.1. Proof of Lemma SM12.1.**

*Proof.* Proof is based on the two following lemmas.

**Lemma SM12.2.** *Let  $\lambda \in \mathbb{R}_+^p$  and  $A$  satisfies (2.3), (2.4). Then, for any compact  $U \subset \text{Span}(A^T)$  it holds that*

$$(SM12.3) \quad S_{A,\lambda}(U) = (\lambda + U + \ker A) \cap \mathbb{R}_+^p \text{ is convex and compact,}$$

where the summation sign denotes the Minkowski sum

$$A + B = \{w = u + v \in \mathbb{R}^p : u \in A, v \in B\}, \quad A \subset \mathbb{R}^p, B \subset \mathbb{R}^p.$$

**Lemma SM12.3.** *Let assumptions of Lemma SM12.2 be satisfied and  $d_H(A, B)$  denote the Hausdorff distance between compact sets  $A, B \subset \mathbb{R}^p$  being defined by the formula*

$$d_H(A, B) = \max \left( \sup_{x \in A} \inf_{y \in B} \|x - y\|, \sup_{x \in B} \inf_{y \in A} \|x - y\| \right).$$

Let  $U \subset \text{Span}(A^T)$  be a compact such that  $S_{A,\lambda}(U) \neq \emptyset$ . Then,

$$(SM12.4) \quad d_H(S_{A,\lambda}(\{u_0\}), S_{A,\lambda}(\{u\})) \rightarrow 0 \text{ for } u \rightarrow u_0, u, u_0 \in U,$$

where  $S_{A,\lambda}(\cdot)$  is defined in (SM12.3).

From the result of Lemma SM12.2 and the assumption in (2.12) it follows that for each  $u \in U$  the following problem

$$(SM12.5) \quad \begin{aligned} & \text{minimize } \varphi(\lambda + u + w) \text{ w.r.t } w, \\ & \text{subject to: } \lambda + u + w \succeq 0, w \in \ker A, \end{aligned}$$

admits a unique solution  $w(u) \in \ker A$ . Indeed, the minimized function in (SM12.5) is strictly convex function in  $w$  and the domain is compact and convex. This proves the first assertion of the lemma.

Now, we prove the continuity of  $w(u)$  on its domain. Let  $u_k$  be a sequence in  $U$  such that  $u_k \rightarrow u_0$  for some  $u_0 \in U$ . Let  $w_k = w(u_k)$ , where the latter are minimizers in (SM12.5) for  $u = u_k$ , and  $w_0 = w(u_0)$ . We know that  $\lambda_k = \lambda + u_k + w(u_k) \in S_{A,\lambda}(U)$ , where the latter is a compact (by Lemma SM12.2). Since continuous mapping of a compact is again a compact, all  $w_k$  belong to some compact  $W_{A,\lambda}(U)$  being the orthogonal projection of  $(S_{A,\lambda}(U) - \lambda)$  onto  $\ker A$ . From compactness of  $W_{A,\lambda}(U)$  it follows that  $w_k$  contains a converging subsequence  $w_m \rightarrow w_0$ ,  $w_0 \in W_{A,\lambda}(U)$ , where  $w_m = w(u_m)$ ,  $m \in \mathbb{N}$ .

Since  $w_m$  are the minimizers in (SM12.5), we know that

$$(SM12.6) \quad \begin{aligned} & \varphi(\lambda + u_m + w_m) \leq \varphi(\lambda + u_m + w), \\ & \text{for all } w \in \ker A, \text{ such that } \lambda + u_m + w \succeq 0. \end{aligned}$$

Taking the limit  $m \rightarrow +\infty$ ,  $u_m \rightarrow u_0$ ,  $w_m \rightarrow w_0$  we aim to show that

$$(SM12.7) \quad \begin{aligned} \varphi(\lambda + u_0 + w_0) &\leq \varphi(\lambda + u_0 + w), \\ \text{for all } w \in \ker A, \text{ such that } \lambda + u_0 + w &\succeq 0. \end{aligned}$$

Therefore,  $w_0 = w(u_0)$  which is unique (by the strict convexity of  $\varphi$  along  $\ker A$ ) and proves the continuity of  $w(u)$ . The fact that any sequence has a convergent subsequence having the same limit  $w(u_0)$  implies that  $w_k = w(u_k)$  also converges to  $w(u_0)$ . However, taking the limit  $m \rightarrow +\infty$  for each  $w$  in (SM12.6) may not preserve the positivity constraint. To show (SM12.7), for each  $w$  satisfying the positivity constraint in (SM12.7) we find another sequence  $\{w'_m\}$  such that

$$(SM12.8) \quad \lambda + u_m + w'_m \succeq 0, w'_m \rightarrow w \text{ for } m \rightarrow +\infty.$$

In this case we can replace  $w$  with  $w'_m$  in (SM12.6) and take the limit  $m \rightarrow \infty$  in order to obtain (SM12.7).

Now, it is left how to choose  $w'_m$  so that (SM12.8) holds. We choose  $w'_m$  to be the solution in the following minimization problem

$$(SM12.9) \quad \begin{aligned} \text{minimize } &\|(\lambda + u_0 + w) - (\lambda + u_m + w'_m)\| \text{ with respect to } w'_m, \\ \text{subject to: } &w'_m \in \ker A, \lambda + u_m + w'_m \succeq 0. \end{aligned}$$

Solution  $w'_m$  in (SM12.9) always exists and unique since it corresponds to the euclidean projection of  $\lambda + u_0 + w$  onto convex set  $S_{A,\lambda}(\{u_m\})$ , that is

$$(SM12.10) \quad w'_m = \Pi_{\ker A}[\text{Proj}(\lambda + u_0 + w, S_{A,\lambda}(\{u_m\})) - \lambda],$$

where  $\Pi_{\ker A}$  is the orthogonal projector onto  $\ker A$ ,  $\text{Proj}(x, X)$  denotes the euclidean projection of point  $x$  onto  $X$ . From (SM12.10) and the fact that  $\lambda + u_0 + w \in S_{A,\lambda}(\{u_0\})$  it follows that

$$(SM12.11) \quad w'_m - w = \Pi_{\ker A}[\text{Proj}(\lambda + u_0 + w, S_{A,\lambda}(\{u_m\})) - \text{Proj}(\lambda + u_0 + w, S_{A,\lambda}(\{u_0\}))].$$

Using (SM12.11) and Proposition 5.3 from [SM1] one can write the following estimate:

$$(SM12.12) \quad \|w'_m - w\| \leq \rho_m^{1/2} d_{H,\rho_m}(S_{A,\lambda}(\{u_0\}), S_{A,\lambda}(\{u_m\}))^{1/2},$$

where  $\rho_m = \|\lambda + u_0 + w\| + d(\lambda + u_0 + w, S_{A,\lambda}(\{u_m\}))$  ( $d(x, y)$  denotes the standard euclidean distance between  $x, y$ ,  $d(x, X) = \inf_{x' \in X} d(x, x')$ ),  $d_{H,\rho}(\cdot, \cdot)$  is the bounded Hausdorff distance (see the definition in Section 3 of [SM1]). In particular, for  $d_{H,\rho}$  the following bound holds:

$$(SM12.13) \quad d_{H,\rho}(A, B) \leq d_H(A, B),$$

for any sets  $A, B$ .

First, note that  $\sup_m \rho_m$  is finite. Indeed, this follows from the fact that  $u_m \rightarrow u_0$  (hence  $\{u_m\}$  is bounded) and following estimates:

$$(SM12.14) \quad \begin{aligned} d(\lambda + u_0 + w, S_{A,\lambda}(\{u_m\})) &\leq d(\lambda + u_0 + w, 0) + d(0, S_{A,\lambda}(\{u_m\})) \\ &\leq \|\lambda + u_0 + w\| + d(0, S_{A,\lambda}(\{u_m\})), \end{aligned}$$

$$(SM12.15) \quad d(0, S_{A,\lambda_*}(\{u_m\})) \leq \max_{j \in \{1, \dots, p\}} \left( \sum_{i=1}^d a_i^T(\lambda + u_m) \right) / A_j, \quad A_j = \sum_{i=1}^d a_{ij}.$$

Formula (SM12.14) is a simple triangle inequality and the estimate in (SM12.15) follows from the fact that  $S_{A,\lambda}(\{u\})$  is the affine subset of  $(p-1)$ -simplex defined by the formula

$$(SM12.16) \quad \Delta_{A,\lambda}^p(u) = \{\lambda' \in \mathbb{R}_+^p : \sum_{j=1}^p \lambda'_j A_j = \sum_{i=1}^d a_i^T(\lambda + u) \geq 0\}, \quad A_j = \sum_{i=1}^d a_{ij} > 0.$$

So the inequality in (SM12.15) express the fact that the furthest point from the origin to  $\Delta_{A,\lambda}^p$  is one of its vertices. From (SM12.12), (SM12.13) the fact that  $\sup_m \rho_m < +\infty$  and the result of Lemma SM12.3 it follows that  $w'_m \rightarrow w$ , where  $\lambda + u_m + w_m \geq 0$ . Therefore, conditions in (SM12.8) are satisfied which, in turn, proves (SM12.7) and the second claim of the lemma.

Lemma is proved. ■

### SM12.2. Proof of Lemma SM12.2.

*Proof.* Closedness and convexity of  $S_{A,\lambda}(U)$  follow directly from the fact that  $(\lambda + U + \ker A)$ ,  $\mathbb{R}_+^p$  are both closed and convex whereas their intersection preserves these properties.

We prove boundedness of  $S_{A,\lambda}(U)$  by the contradiction argument.

Assume that  $S_{A,\lambda}(U)$  is not bounded, then there exists a sequence  $\{(u_k, w_k)\}_{k=1}^\infty$ ,  $u_k \in U$ ,  $w_k \in \ker A$ , such that

$$(SM12.17) \quad \lambda + u_k + w_k \in \mathbb{R}_+^p, \quad \|\lambda + u_k + w_k\| \rightarrow \infty.$$

From (SM12.17) and compactness of  $U$  it follows, in particular, that

$$(SM12.18) \quad w_k \in \ker A, \quad \|w_k\| \rightarrow +\infty.$$

Also there exists a converging subsequence  $\{u_{k_n}\}_{n=1}^\infty$  such that

$$(SM12.19) \quad u_{k_n} \rightarrow u_0 \in U \text{ for some } u_0, \text{ as } n \rightarrow +\infty.$$

Consider the corresponding subsequence  $\{w_{k_n}\}_{n=1}^\infty$  for which we know that

$$(SM12.20) \quad w_{k_n} \in \ker A, \quad \|w_{k_n}\| \rightarrow +\infty \text{ for } n \rightarrow +\infty.$$

Let

$$(SM12.21) \quad \theta_n = \frac{w_{k_n}}{\|w_{k_n}\|}, \quad \theta_n \in \mathbb{S}^{p-1} \cap \ker A.$$

Since  $\mathbb{S}^{p-1} \cap \ker A$  is compact,  $\{\theta_n\}_{n=1}^\infty$  has a converging subsequence  $\{\theta_m\}_{m=1}^\infty$  such that

$$(SM12.22) \quad \theta_m \rightarrow \theta_0, \theta_0 \in \mathbb{S}^{p-1} \cap \ker A.$$

Let  $\{u_m\}_{m=1}^\infty$  be the corresponding subsequence of  $\{u_{k_n}\}_{n=1}^\infty$  for index  $m$  in formula (SM12.22). From (SM12.17)-(SM12.22) it follows that we have constructed a sequence  $\{(u_m, w_m)\}_{m=1}^\infty$  such that

$$(SM12.23) \quad \lambda + u_m + w_m \in \mathbb{R}_+^p, u_m \in U, w_m \in \ker A,$$

$$(SM12.24) \quad u_m \rightarrow u_0, \|w_m\| \rightarrow +\infty,$$

$$(SM12.25) \quad \theta_m = \frac{w_m}{\|w_m\|} \rightarrow \theta_0 \in \mathbb{S}^{p-1} \cap \ker A.$$

Now we show that under our initial assumption we arrive to the fact that

$$(SM12.26) \quad \lambda + s\theta_0 \in \mathbb{R}_+^p \text{ for any } s > 0,$$

where  $\theta_0$  is defined in (SM12.25).

Indeed, from the fact that  $\lambda \in \mathbb{R}_+^p$  and that  $\mathbb{R}_+^p$  is convex it follows that

$$(SM12.27) \quad \lambda + t(u_m + w_m) = \lambda + t(u_m + \|w_m\|\theta_m) \in \mathbb{R}_+^p \text{ for any } t \in [0, 1].$$

Let  $s > 0$ . By choosing  $t = t_m(s) = s/\|w_m\|$  in (SM12.27) ( $t_m(s) \in [0, 1]$  for large  $m$ ; see (SM12.24)) and using formulas (SM12.23)–(SM12.25) we obtain

$$(SM12.28) \quad \begin{aligned} & (\lambda + s\theta_0) - (\lambda + t_m(s)u_m + t_m(s)\|w_m\|\theta_m) \\ & = s(\theta_0 - \theta_m) - s \frac{u_m}{\|w_m\|} \rightarrow 0 \text{ for } m \rightarrow +\infty. \end{aligned}$$

From (SM12.28) it follows that  $\lambda + s\theta_0$  is a limiting point in  $\mathbb{R}_+^p$ , and due to its closedness it follows that  $\lambda + s\theta_0 \in \mathbb{R}_+^p$ ,  $s \geq 0$ .

The statement in (SM12.26) cannot hold, because from (2.5) it follows that

$$(SM12.29) \quad \text{for any } \theta \in \ker A, \theta \neq 0 \exists j \in \{1, \dots, p\} \text{ s.t. } \theta_j < 0.$$

Since  $\theta_0 \in \ker A$ , by taking  $s > 0$  large enough in formula (SM12.26), we will arrive to the case when  $\lambda + s\theta_0 \notin \mathbb{R}_+^p$ , which gives the desired contradiction.

Lemma is proved. ■

### SM12.3. Proof of Lemma SM12.3.

*Proof.* The claim of the lemma makes part of Theorem 1 from [SM31] which, informally says that a closed convex set  $K \subset \mathbb{R}^p$  is a polyhedra iff the Hausdorff distance on the space sections by any family of parallel linear subspaces is Lipschitz continuous with respect to the shift vector.

Using notations from [SM31] we define the following affine mapping

$$(SM12.30) \quad \tau_{A,\lambda}(u) = A\lambda + Au, u \in \mathbb{R}^p,$$

where  $\lambda$  is a parameter,  $A \in \text{Mat}(d, p)$  is the design matrix satisfying (2.3) and (2.4).

Let  $K = \mathbb{R}_+^p$  which is obviously a polyhedra in  $\mathbb{R}^p$ . Next, we define family of sections of  $K$  by the formula

$$(SM12.31) \quad k(\Lambda) = \tau_{A,\lambda}^{-1}(\Lambda) \cap K, \quad \Lambda \in \mathbb{R}^d.$$

Essentially,  $k(\Lambda)$  is an section of  $K$  by  $\ker A$  which is shifted by vector  $u$  (in some cases  $k(\Lambda)$  can be an empty set). In particular, if  $\Lambda = \Lambda(u) = A\lambda + Au$  for some  $u \in \text{Span}(A^T)$ , then it is easy to see that

$$(SM12.32) \quad k(\Lambda(u)) = (\lambda + u + \ker A) \cap K = (\lambda + u + \ker A) \cap \mathbb{R}_+^p = S_{A,\lambda}(\{u\}),$$

where  $S_{A,\lambda}$  is defined in (SM12.3).

The result of Theorem 1 from [SM31] says, in particular, that

$$(SM12.33) \quad d_H(k(\Lambda), k(\Lambda')) \leq C \|\Lambda - \Lambda'\|,$$

where  $C$  is some constant depending on  $K$  and  $A$ ,  $d_H(\cdot, \cdot)$  is the standard Hausdorff distance being also extended for empty sets. However, this extension is not needed for us since we always consider parameters  $\Lambda(u)$  for  $u$  from some  $U \subset \text{Span}(A^T)$  with apriori non-empty sets  $S_{A,\lambda}(\{u\})$ .

From formulas (SM12.32), (SM12.33) it follows that

$$(SM12.34) \quad d_H(S_{A,\lambda}(\{u\}), S_{A,\lambda}(\{u'\})) \leq C \|A(u - u')\|,$$

which directly implies (SM12.4).

Lemma is proved. ■

**SM12.4. Proofs of theorems Theorems 6.4 and 6.5.** First we prove Theorem 6.5, then we show that if (6.6) holds conditions in (6.8) for Theorem 6.5 are satisfied which, in turn, automatically proves Theorem 6.4.

*Proof of Theorem 6.5.* Using (2.9), (2.10), the minimization problem in Step 6 of Algorithm 4 can be rewritten as as follows:

$$(SM12.35) \quad \begin{aligned} \tilde{\lambda}_b^t &= \arg \min_{\lambda \geq 0} L_p(\lambda \mid \tilde{\Lambda}_b^t, A, 1, \beta^t/t) \\ &= \arg \min_{\lambda \geq 0} \mathcal{L}^t(\lambda), \end{aligned}$$

where

$$(SM12.36) \quad \begin{aligned} \mathcal{L}^t(\lambda) &= \sum_{i \in I_1(\Lambda^*)} (-\tilde{\Lambda}_{b,i}^t + \Lambda_i^*) \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) \\ &+ \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) + (\Lambda_i - \Lambda_i^*) \\ &+ \sum_{i \in I_0(\Lambda^*)} -\tilde{\Lambda}_{b,i}^t \log(\Lambda_i) + \Lambda_i + \frac{\beta_t}{t} (\varphi(\lambda) - \varphi(\lambda_*)), \end{aligned}$$

where  $I_0(\cdot)$ ,  $I_1(\cdot)$  are defined in (2.2) and  $\Lambda^* = A\lambda_*$ .

Next, for the proof we use the following lemma.

**Lemma SM12.4.** Let  $\mathcal{L}^t(\lambda)$  be defined in (SM12.36) and conditions of Theorem 6.5 be satisfied. Let  $C_{A,\delta}(\lambda')$ ,  $\delta > 0$ ,  $\lambda' \succeq 0$ , be the cylinder set defined by the formula

$$(SM12.37) \quad C_{A,\delta}(\lambda') = \{\lambda \in \mathbb{R}_+^p, \lambda = \lambda' + \delta u + w \mid (u, w) \in \text{Span}(A^T) \times \ker A, \|u\| = 1\}.$$

Then,

i) there exists  $\delta_0 = \delta_0(A, \lambda_*) > 0$  such that for any  $\delta < \delta_0$  it holds that

$$(SM12.38) \quad \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \mathcal{L}^t(\lambda) \geq C\delta^2 + o_{cp}(1) \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty).$$

where  $C$  is a positive constant independent of  $\delta$ .

ii) there exists a family of random variables  $\tilde{\lambda}^t \in \mathbb{R}_+^p$ ,  $t \in [0, +\infty)$ , such that

$$(SM12.39) \quad \tilde{\lambda}^t \xrightarrow{c.p.} \lambda_* \text{ and } \mathcal{L}^t(\tilde{\lambda}^t) \xrightarrow{c.p.} 0.$$

From the result of Lemma SM12.4(i) it follows that for all  $\lambda \succeq 0$  at distance  $\delta$  from  $\lambda_*$  in the  $\text{Span}(A^T)$  values of  $\mathcal{L}^t(\lambda)$  are greater or equal than  $C\delta^2$  with conditional probability tending to one a.s.  $Y^t$ ,  $t \in [0, +\infty)$ . At the same time, result of Lemma SM12.4(ii) says that there is  $\tilde{\lambda}^t \in \mathbb{R}_+^p$  which is arbitrarily close to  $\lambda_*$  and  $\mathcal{L}^t(\tilde{\lambda}^t)$  is converges to zero for  $t \rightarrow +\infty$  with conditional probability also tending to one. The fact that  $\mathcal{L}^t(\lambda)$  is convex together with the above arguments and  $\tilde{\lambda}_b^t$  being the unique minimizer of  $\mathcal{L}^t(\lambda)$  imply that

$$(SM12.40) \quad P(\|\Pi_{A^T}(\tilde{\lambda}_b^t - \lambda_*)\| < \delta \mid Y^t, t) \rightarrow 1 \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty).$$

where  $\Pi_{A^T}$  is the orthogonal projector onto  $\text{Span}(A^T)$ . Since  $\delta$  can be chosen arbitrarily small in Lemma SM12.4 formula (SM12.40) implies that

$$(SM12.41) \quad \Pi_{A^T}(\tilde{\lambda}_b^t - \lambda_*) \xrightarrow{c.p.} 0.$$

Vector  $\tilde{\lambda}_b^t$  admits in a unique way the following representation

$$(SM12.42) \quad \tilde{\lambda}_b^t = \lambda_* + \tilde{u}_b^t + \tilde{w}_b^t, \text{ where } (\tilde{u}_b^t, \tilde{w}_b^t) \in \text{Span}(A^T) \times \ker A.$$

Using (SM12.35), (SM12.36), and (SM12.42) one can see that

$$(SM12.43) \quad \tilde{w}_b^t = \arg \min_{\substack{w: \lambda_* + \tilde{u}_b^t + w \succeq 0, \\ w \in \ker A}} \varphi(\lambda_* + \tilde{u}_b^t + w) = w_{A,\lambda_*}(\tilde{u}_b^t),$$

where  $w_{A,\lambda}(\cdot)$  is defined in (SM12.2). From (SM12.43), the fact that  $\tilde{u}_b^t \xrightarrow{c.p.} 0$  (see formulas (SM12.41), (SM12.42)), continuity of the map  $w_{A,\lambda_*}(\cdot)$  (by the result of Lemma SM12.1) and the Continuous Mapping Theorem (see, e.g. [SM29], Theorem 2.3, p. 7) it follows that

$$(SM12.44) \quad \tilde{w}_b^t \xrightarrow{c.p.} w_{A,\lambda_*}(0).$$

Formula (6.7) follows directly from (SM12.41)–(SM12.44) and the definition of  $\lambda_{*opt}$ .

Theorem is proved. ■

*Proof of Theorem 6.4.* To prove the theorem we use the following lemma.

**Lemma SM12.5.** *Let  $\tilde{\lambda}_b^t$  be defined as in Algorithm 4 and let  $\theta^t/t \rightarrow 0$  when  $t \rightarrow +\infty$ . Then,*

$$(SM12.45) \quad \tilde{\Lambda}_{b,i}^t \xrightarrow{c.p.} \Lambda_i^* = a_i^T \lambda_*.$$

In view of (SM12.45) in Lemma SM12.5 all assumptions for Theorem 6.5 are satisfied, which implies formula (6.7).

Theorem is proved. ■

### SM12.5. Proof of Lemma SM12.4.

*Proof.* First we prove (i), then for (ii) we give an explicit formula for  $\tilde{\lambda}^t$  for which (SM12.39) holds.

First, in formula (SM12.36) one can see that

$$(SM12.46) \quad \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \sum_{i \in I_1(\Lambda^*)} \left( -\tilde{\Lambda}_{b,i}^t + \Lambda_i^* \right) \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) \xrightarrow{c.p.} 0.$$

The above formula follows from the assumption that  $\tilde{\Lambda}_{b,i}^t \xrightarrow{c.p.} \Lambda_i^*$  and that  $\log(\Lambda_i/\Lambda_i^*) = \log(1 + \delta a_i^T u / \Lambda_i^*)$  is uniformly bounded for  $\lambda \in C_{A,\delta}(\lambda_*)$  from above and below for  $\delta$  small enough ( $u \in \text{Span}(A^T)$ ,  $\|u\| = 1$ ). For example, to bound all of the logarithmic terms in (SM12.46) we may choose any  $\delta$  such that

$$(SM12.47) \quad 0 < \delta < \min_{i \in I_1(\Lambda^*)} (\Lambda_i^* \|a_i\|^{-1}).$$

Since  $\varphi(\lambda)$  satisfies (2.11), (2.12), there exists a constant  $M = M(\lambda_*, \delta, A)$  such that

$$(SM12.48) \quad \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \varphi(\lambda) \geq M.$$

From (6.6) and (SM12.48) it follows that

$$(SM12.49) \quad (\beta^t/t) \inf_{C_{A,\delta}(\lambda_*)} (\varphi(\lambda) - \varphi(\lambda_*)) \geq o(1), \text{ when } t \rightarrow +\infty.$$

Using (SM12.36), (SM12.46), and (SM12.49) we obtain the following estimate

$$(SM12.50) \quad \begin{aligned} \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \mathcal{L}^t(\lambda) &\geq o_{cp}(1) + \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) + (\Lambda_i - \Lambda_i^*) \\ &+ \sum_{i \in I_0(\Lambda^*)} -\tilde{\Lambda}_{b,i}^t \log(\Lambda_i) + \Lambda_i. \end{aligned}$$

Note that

$$(SM12.51) \quad -\tilde{\Lambda}_{b,i}^t \log(\Lambda_i) \geq 0 \text{ for } \Lambda_i \leq 1, i \in I_0(\Lambda^*).$$



From (2.2), (SM12.37) it follows that we can choose  $\delta$  sufficiently small so that

$$(SM12.52) \quad \Lambda_i \leq 1 \text{ for all } \lambda \in C_{A,\delta}(\lambda_*), i \in I_0(\Lambda^*).$$

For example, it suffices to choose  $\delta$  as follows

$$(SM12.53) \quad 0 < \delta \leq \min_{i \in \{1, \dots, d\}} (\|a_i\|^{-1}).$$

Using (SM12.50), (SM12.51), for  $\delta$  satisfying (SM12.47) and (SM12.53) we obtain

$$(SM12.54) \quad \inf_{\lambda \in C_{A,\delta}(\Lambda^*)} \mathcal{L}^t(\lambda) \geq \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) + (\Lambda_i - \Lambda_i^*) \\ + \sum_{i \in I_0(\Lambda^*)} \Lambda_i + o_{cp}(1).$$

Now consider

$$(SM12.55) \quad \Phi_{s^*}(s) = -s^* \log(s) + s, \quad s > 0, \quad s^* > 0.$$

Function  $\Phi_{s^*}(s)$  is convex, smooth, has positive non-vanishing second derivative  $\Phi_{s^*}''(s)$  and at  $s = s^*$  it has its global minimum. Therefore, for any  $\varepsilon > 0$  small enough (for example, for  $\varepsilon < s^*$ ) there exists positive constant  $C(\varepsilon, s^*)$  such that

$$(SM12.56) \quad \Phi_{s^*}(s) - \Phi_{s^*}(s^*) \geq C(\varepsilon, s^*) |s - s^*|^2 \text{ for } |s - s^*| < \varepsilon.$$

From (SM12.56) it follows that one can choose  $\delta_0 > 0$  such that

$$(SM12.57) \quad \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) + (\Lambda_i - \Lambda_i^*) \\ + \sum_{i \in I_0(\Lambda^*)} \Lambda_i \geq C(\delta_0, \Lambda^*) \sum_{i \in I_1(\Lambda^*)} (\Lambda_i - \Lambda_i^*)^2 + \sum_{i \in I_0(\Lambda^*)} \Lambda_i \\ \text{for } |\Lambda_i - \Lambda_i^*| < \delta_0, i \in I_1(\Lambda^*).$$

Value for  $\delta_0$  is precised below. Let  $\lambda \in C_{A,\delta}(\lambda_*)$  and  $\delta < \delta_0$ , that is  $\lambda = \lambda_* + \delta u + w$ , where  $u \in \text{Span}(A^T)$ ,  $\|u\| = 1$ ,  $w \in \ker A$ . For  $\delta$  satisfying (SM12.53) formula (SM12.52) holds and we get the following estimate:

$$(SM12.58) \quad \Lambda_i = a_i^T \lambda = \delta a_i^T u \geq \delta^2 (a_i^T u)^2 \geq 0 \text{ for } i \in I_0(\Lambda^*).$$

In (SM12.58) we used the fact that  $\Lambda_i^* = a_i^T \lambda_* = 0$ ,  $i \in I_0(\Lambda^*)$ .

From (SM12.57), (SM12.58) it follows that

$$(SM12.59) \quad \inf_{\lambda \in C_{A,\delta}(\lambda_*)} \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \log \left( \frac{\Lambda_i}{\Lambda_i^*} \right) + (\Lambda_i - \Lambda_i^*) + \sum_{i \in I_0(\Lambda^*)} \Lambda_i \\ \geq \min(C(\delta_0, \Lambda^*), 1) \delta^2 \sum_{i=1}^d (a_i^T u)^2 \\ \geq \min(C(\delta_0, \Lambda^*), 1) \delta^2 \sigma_{\min}^+(A^T A),$$

where  $\sigma_{min}^+(A^T A)$  is the smallest non-zero eigenvalue of  $A^T A$  on the  $\text{Span}(A^T)$ . In particular, in (SM12.58), (SM12.59) we have used the property that  $u \in \text{Span}(A^T)$  which guarantees that

$$(SM12.60) \quad \sum_{i=1}^d (a_i^T u)^2 = u^T A^T A u \geq \sigma_{min}^+(A^T A) > 0 \text{ for } \|u\| = 1.$$

Formula (SM12.38) follows directly from (SM12.54) and (SM12.59).

Finally, we choose  $\delta_0$  as follows

$$(SM12.61) \quad \delta_0 = \frac{1}{2} \min \left[ \min_{i \in \{1, \dots, d\}} (\|a_i\|^{-1}), \min_{i \in I_1(\Lambda^*)} (\Lambda_i^* \|a_i\|^{-1}), \min_{i \in I_1(\Lambda^*)} \Lambda_i^* \right],$$

so that conditions (SM12.47), (SM12.53) are simultaneously satisfied together with (SM12.57).

Part (i) of Lemma SM12.4 is proved. Now we prove part (ii) of the lemma.

Let

$$(SM12.62) \quad \tilde{\lambda}^t = \lambda_* + \sum_{i \in I_0(\Lambda^*)} \tilde{\Lambda}_{b,i}^t \frac{a_i}{\|a_i\|^2}.$$

Note that  $\tilde{\lambda}^t \in \mathbb{R}_+^p$  because  $a_i \in \mathbb{R}_+^p$  and  $\tilde{\Lambda}_{b,i}^t \geq 0$ . Since  $\tilde{\Lambda}_{b,i}^t \xrightarrow{c.p.} 0$  for  $i \in I_0(\Lambda^*)$  (by the assumption) we immediately have that

$$(SM12.63) \quad \tilde{\lambda}^t \xrightarrow{c.p.} \lambda_*.$$

Note that in (SM12.36) for  $\mathcal{L}^t(\lambda)$  all summands are continuous and equal to zero at  $\lambda = \lambda_*$  except the logarithmic part

$$(SM12.64) \quad g(\lambda) = \sum_{i \in I_0(\Lambda^*)} -\tilde{\Lambda}_{b,i}^t \log(\Lambda_i), \quad \Lambda_i = a_i^T \lambda.$$

From the fact that  $a_i \in \mathbb{R}_+^p$  (see formula (2.3)) it follows that  $a_i^T a_{i'} \geq 0$  for all  $i, i'$ . Using this property and monotonicity of the logarithm ( $\log(x+y) \geq \log(x)$  for  $y \geq 0$ ) it follows that

$$(SM12.65) \quad \begin{aligned} g(\tilde{\lambda}^t) &= \sum_{i \in I_0(\Lambda^*)} -\tilde{\Lambda}_{b,i}^t \log(a_i^T \tilde{\lambda}^t) \\ &\leq \sum_{i \in I_0(\Lambda^*)} -\tilde{\Lambda}_{b,i}^t \log(\tilde{\Lambda}_{b,i}^t) \xrightarrow{c.p.} 0. \end{aligned}$$

Formula (SM12.65) gives an asymptotic upper bound on  $g(\tilde{\lambda}^t)$  which is equal to zero. For the lower bound we use formulas (SM12.51), (SM12.63) and the fact that  $a_i^T \tilde{\lambda}^t \xrightarrow{c.p.} 0$  for  $i \in I_0(\Lambda^*)$  from which it follows that

$$(SM12.66) \quad \begin{aligned} g(\tilde{\lambda}^t) &\geq 0 \text{ with conditional probability tending to one for } t \rightarrow +\infty, \\ &\text{a.s. } Y^t, t \in [0, +\infty). \end{aligned}$$

From (SM12.65), (SM12.66) it follows that

$$(SM12.67) \quad g(\tilde{\lambda}^t) \xrightarrow{c.p.} 0.$$

From (SM12.36), (SM12.62), (SM12.64), and (SM12.67) it follows that

$$(SM12.68) \quad \mathcal{L}^t(\tilde{\lambda}^t) \xrightarrow{c.p.} 0.$$

This proves part (ii) of the lemma.

Lemma is proved. ■

### SM12.6. Proof of Lemma SM12.5.

*Proof.* Recall that

$$(SM12.69) \quad \tilde{\Lambda}_{b,i}^t \mid Y^t, \tilde{\Lambda}_{\mathcal{M}}^t, t \sim \Gamma(Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t, (\theta^t + t)^{-1}), \quad i \in \{1, \dots, d\},$$

where  $\tilde{\Lambda}_{\mathcal{M}}^t \mid Y^t, t$  is sampled in Algorithm 3. From the definition of  $\tilde{\Lambda}^t$  in Step 3 of Algorithm 3 and necessary optimality conditions in Step 4 (see also analogous formula (SM12.78)) it follows that

$$(SM12.70) \quad \sum_{i=1}^d \tilde{\Lambda}_{\mathcal{M},i}^t = \sum_{i=1}^d \tilde{\Lambda}_i^t,$$

$$(SM12.71) \quad \tilde{\Lambda}_{\mathcal{M}}^t \succeq 0, \tilde{\Lambda}^t \succeq 0, E[\tilde{\Lambda}_i^t \mid Y^t, t] = Y_i^t/t, \quad i \in 1, \dots, d.$$

Using (SM12.70), (SM12.71) we get the following estimate:

$$(SM12.72) \quad E[\tilde{\Lambda}_{\mathcal{M},i}^t \mid Y^t, t] \leq \sum_{i=1}^d \frac{Y_i^t}{t}, \quad i \in \{1, \dots, d\}.$$

Let  $\varepsilon > 0$ . Using the Markov inequality we obtain

$$(SM12.73) \quad \begin{aligned} p(|\tilde{\Lambda}_{b,i}^t - \Lambda_i^*| > \varepsilon \mid Y^t, t) &\leq \frac{E[|\tilde{\Lambda}_{b,i}^t - \Lambda_i^*| \mid Y^t, t]}{\varepsilon} \\ &\leq \frac{E[|\tilde{\Lambda}_{b,i}^t - \frac{Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t}| \mid Y^t, t]}{\varepsilon} + \frac{E[|\frac{Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t} - \Lambda_i^*| \mid Y^t, t]}{\varepsilon}. \end{aligned}$$

Using the Jensen's inequality  $E|X|^2 \geq (E|X|)^2$ , formulas (SM12.69), (SM12.72), the Strong Law of Large Numbers for  $Y^t$  (see Theorem SM2.1(i) in section SM2) and the fact that

$\theta^t/t \rightarrow 0$ , we get the following:

(SM12.74)

$$\begin{aligned} E\left[\left|\tilde{\Lambda}_{b,i}^t - \frac{Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t}\right| \mid Y^t, t\right] &\leq \left(E\left[\left|\tilde{\Lambda}_{b,i}^t - \frac{Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t}\right|^2 \mid Y^t, t\right]\right)^{1/2} \\ &= \left(E[\text{var}[(\tilde{\Lambda}_{b,i}^t) \mid Y^t, \tilde{\Lambda}_{\mathcal{M},i}^t, t] \mid Y^t, t]\right)^{1/2} \\ &= \left(\frac{Y_i^t + \theta^t E[\tilde{\Lambda}_{\mathcal{M},i}^t \mid Y^t, t]}{(t + \theta^t)^2}\right)^{1/2} \\ &\leq \left(\frac{Y_i^t + (\theta^t/t) \sum_{i=1}^d Y_i^t}{(t + \theta^t)^2}\right)^{1/2} \rightarrow 0 \text{ a.s. } Y^t, t \in [0, +\infty). \end{aligned}$$

For estimation of the second term in (SM12.73) we use formula (SM12.72), the triangle inequality and again the property that  $\theta^t/t \rightarrow 0$  to get the following:

(SM12.75)

$$\begin{aligned} E\left[\left|\frac{Y_i^t + \theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t} - \Lambda_i^*\right| \mid Y^t, t\right] &\leq \left|\frac{Y_i^t}{\theta^t + t} - \Lambda_i^*\right| + E\left[\left|\frac{\theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t}\right| \mid Y^t, t\right] \\ &\leq \left|\frac{Y_i^t}{\theta^t + t} - \Lambda_i^*\right| + \frac{\theta^t}{\theta^t + t} \sum_{i=1}^d \frac{Y_i^t}{t} \rightarrow 0 \text{ a.s. } Y^t, t \in [0, +\infty). \end{aligned}$$

Formula (SM12.45) follows from formulas (SM12.73)–(SM12.75).

Lemma is proved. ■

### SM12.7. Proof of Proposition 6.8.

*Proof.* First prove that the set of minimizers in (6.9) is always nonempty and is a subset of the simplex in (6.11). From the Karush-Kuhn-Tucker optimality conditions (see e.g., [SM3], Section 3.3) it follows that

$\exists(\lambda_{\mathcal{M},*}, \mu_{\mathcal{M},*}) \in \mathbb{R}_+^p \times \mathbb{R}_+^p$  such that

$$(SM12.76) \quad \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \frac{a_{\mathcal{M},ij}}{\Lambda_{\mathcal{M},i}^*} + \sum_{i=1}^d a_{\mathcal{M},ij} - \mu_{\mathcal{M},*,j} = 0,$$

$$(SM12.77) \quad \mu_{\mathcal{M},*,j} \lambda_{\mathcal{M},*,j} \equiv 0, \text{ for all } j \in \{1, \dots, p_{\mathcal{M}}\}.$$

By multiplying both sides of (SM12.76) on  $\lambda_{\mathcal{M},*,j}$ , summing up all equations with respect to  $j$  and using (SM12.77) we obtain the following necessary optimality condition:

$$(SM12.78) \quad \left\langle \sum_{i=1}^d a_{\mathcal{M},i}, \lambda_{\mathcal{M},*} \right\rangle = \sum_{j=1}^{p_{\mathcal{M}}} A_{\mathcal{M},j} \lambda_{\mathcal{M},*,j} = \sum_{i=1}^d \Lambda_i^*,$$

$$A_{\mathcal{M},j} = \sum_{i=1}^d a_{\mathcal{M},ij}.$$

Formula (SM12.78) proves (6.11). The constraint in (SM12.78) can be added to the set of constraints in (6.9) without any effect since it is necessary. Because the minimized functional in (6.9) is convex and the domain of constraints is now a convex compact there always exists at least one minimizer.

Demonstration of (6.12) is straightforward. Indeed, if for some  $i$  we have  $\Lambda_i^* > 0$ , then necessarily  $\Lambda_{\mathcal{M},i}^* > 0$ , otherwise the value of the target functional becomes  $+\infty$  due to explosion of the logarithmic term. At the same time any interior point  $\lambda_{\mathcal{M}} \in \Delta_{A_{\mathcal{M}}}^{p_{\mathcal{M}}}(\Lambda^*)$  (i.e.,  $\lambda_{\mathcal{M}} \succ 0$ ) would result in the finite value of the target functional. Hence, inclusions (6.12) always hold.

Proposition is proved. ■

### SM12.8. Proof of Theorem 6.9.

*Proof.* First we prove (6.13), then (6.15) which also implies uniqueness of the minimizer. Let  $\lambda_{\mathcal{M},*} \in \mathbb{R}_+^{p_{\mathcal{M}}}$  be a minimizer in (6.9) (possibly not unique; see also Proposition 6.8). Let

$$(SM12.79) \quad \lambda_{\mathcal{M}} = \lambda_{\mathcal{M},*} + u_{\mathcal{M}}, \quad \lambda_{\mathcal{M}} \in \mathbb{R}_+^{p_{\mathcal{M}}}.$$

Consider the second order Taylor expansion of  $L(\lambda \mid \Lambda^*, A_{\mathcal{M}}, 1)$  in (6.9) in a vicinity of  $\lambda_{\mathcal{M},*}$ :

$$(SM12.80) \quad \begin{aligned} & L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1) - L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) \\ &= u_{\mathcal{M}}^T \nabla L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) + \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \Lambda_i^* \frac{(u_{\mathcal{M}}^T a_{\mathcal{M},i})^2}{(\Lambda_{\mathcal{M},i}^*)^2} \\ &+ o(\|\Pi_{A_{\mathcal{M},I_1(\Lambda^*)}^T} u_{\mathcal{M}}\|^2), \end{aligned}$$

where  $\Lambda_{\mathcal{M}}^* = A_{\mathcal{M}} \lambda_{\mathcal{M},*}$  and

$$(SM12.81) \quad \nabla L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) = \sum_{i \in I_1(\Lambda^*)} -\Lambda_i^* \frac{a_{\mathcal{M},i}}{\Lambda_{\mathcal{M},i}^*} + \sum_{i=1}^d a_{\mathcal{M},i}.$$

Karush-Kuhn-Tucker necessary optimality conditions for the problem in (6.9) imply that there exists  $\mu_{\mathcal{M},*}$  such that

$$(SM12.82) \quad \mu_{\mathcal{M},*} \succeq 0, \quad \nabla L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) = \mu_{\mathcal{M},*}, \quad \mu_{\mathcal{M},*,j} \lambda_{\mathcal{M},*,j} = 0, \quad j = 1, \dots, p.$$

From formulas (SM12.79) and (SM12.82) it follows that

$$(SM12.83) \quad \begin{aligned} u_{\mathcal{M}}^T \nabla L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}) &= u_{\mathcal{M}}^T \mu_{\mathcal{M},*} = (\lambda_{\mathcal{M}} - \lambda_{\mathcal{M},*})^T \mu_{\mathcal{M},*} \\ &= \lambda_{\mathcal{M}}^T \mu_{\mathcal{M},*} \geq 0. \end{aligned}$$

Note also that  $\mu_{\mathcal{M},*}$  is the optimal Lagrangian multiplier for the problem in (6.9) for which the strong duality holds (e.g., via Slater condition).

Formulas (6.13) and (6.14) follow from (SM12.80)–(SM12.83). Next, we prove that (6.15) holds.

Using (SM12.83) we obtain the following estimate:

$$(SM12.84) \quad u_{\mathcal{M}}^T \nabla L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) = u_{\mathcal{M}}^T \mu_{\mathcal{M},*} \geq (u_{\mathcal{M}}^T \mu_{\mathcal{M},*})^2 \text{ if } \|u_{\mathcal{M}}\| \leq \|\mu_{\mathcal{M},*}\|^{-1}.$$

From (SM12.80) and (SM12.84) it follows that

$$(SM12.85) \quad \begin{aligned} L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1) - L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) \\ \geq u_{\mathcal{M}}^T C_{\mathcal{M},*} u_{\mathcal{M}} + o(\|u_{\mathcal{M}}\|^2), \\ \text{for } \|u_{\mathcal{M}}\| \leq \|\mu_{\mathcal{M},*}\|^{-1}, \end{aligned}$$

where

$$(SM12.86) \quad C_{\mathcal{M},*} = \mu_{\mathcal{M},*} \mu_{\mathcal{M},*}^T + \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \Lambda_i^* \frac{a_{\mathcal{M},i} a_{\mathcal{M},i}^T}{(\Lambda_{\mathcal{M},i}^*)^2}.$$

To finish the proof we use two following lemmas.

**Lemma SM12.6.** *Let assumptions of Theorem 6.9 be satisfied. Let*

$$(SM12.87) \quad C_{\delta} = \inf_{\substack{u_{\mathcal{M}}: \lambda_{\mathcal{M},*} + u_{\mathcal{M}} \succeq 0, \\ \|u_{\mathcal{M}}\| = \delta}} u_{\mathcal{M}}^T C_{\mathcal{M},*} u_{\mathcal{M}}.$$

Then,

$$(SM12.88) \quad C_{\delta} > 0 \text{ for any } \delta > 0.$$

**Lemma SM12.7.** *Let  $\lambda_{\mathcal{M},*} \in \mathbb{R}_+^{p_{\mathcal{M}}}$ . There exists  $\delta_* > 0$  such that for any  $u_{\mathcal{M}} \in \mathbb{R}^{p_{\mathcal{M}}}$ ,  $0 < |u_{\mathcal{M}}| \leq \delta_*$ ,  $\lambda_{\mathcal{M},*} + u_{\mathcal{M}} \succeq 0$  it also holds that*

$$(SM12.89) \quad \lambda_{\mathcal{M},*} + \delta_* \frac{u_{\mathcal{M}}}{\|u_{\mathcal{M}}\|} \succeq 0.$$

Let  $\delta_*$  be the one of Lemma SM12.7 for chosen  $\lambda_{\mathcal{M},*}$ . From (SM12.85) and (SM12.86) and the results of Lemmas SM12.6 and SM12.7, it follows that

$$(SM12.90) \quad \begin{aligned} L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1) - L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1) \\ \geq \frac{\delta_* u_{\mathcal{M}}^T}{\|u_{\mathcal{M}}\|} C_{\mathcal{M},*} \frac{\delta_* u_{\mathcal{M}}}{\|u_{\mathcal{M}}\|} \frac{\|u_{\mathcal{M}}\|^2}{\delta_*^2} + o(\|u_{\mathcal{M}}\|^2) \\ \geq C_{\delta_*} \frac{\|u_{\mathcal{M}}\|^2}{\delta_*^2} + o(\|u_{\mathcal{M}}\|^2), \quad C_{\delta_*} > 0, \\ \text{for } \lambda_{\mathcal{M}} = \lambda_{\mathcal{M},*} + u_{\mathcal{M}} \succeq 0, |u_{\mathcal{M}}| \leq \min(\delta_*, \|\mu_{\mathcal{M},*}\|^{-1}). \end{aligned}$$

Formula (SM12.90) proves the claim in (6.15).

Theorem is proved. ■

*Proof of Lemma SM12.6.* We use the contradiction argument. Assume that it exists  $\delta > 0$  such that  $C_\delta = 0$ , where  $C_\delta$  is defined in (SM12.87). Since the infimum in (SM12.87) is taken over a compact set, there should exist  $u_{\mathcal{M}}$  such that

$$(SM12.91) \quad \|u_{\mathcal{M}}\| = \delta, \lambda_{\mathcal{M},*} + u_{\mathcal{M}} \succeq 0, u_{\mathcal{M}}^T C_{\mathcal{M},*} u_{\mathcal{M}} = 0.$$

Formulas (SM12.86) and (SM12.91) imply that

$$(SM12.92) \quad u_{\mathcal{M}}^T a_{\mathcal{M},i} = 0, i \in I_1(\Lambda^*), u_{\mathcal{M}}^T \mu_{\mathcal{M},*} = 0.$$

Using (6.10) in the non-expansiveness condition, (SM12.81), (SM12.83), and (SM12.92) we obtain the following:

$$(SM12.93) \quad \begin{aligned} u_{\mathcal{M}}^T \mu_{\mathcal{M},*} &= \sum_{i \in I_0(\Lambda^*)} u_{\mathcal{M}}^T a_{\mathcal{M},*,i} = \sum_{i \in I_0(\Lambda^*)} (\lambda_{\mathcal{M},i} - \lambda_{\mathcal{M},*,i})^T a_{\mathcal{M},*,i} \\ &= \sum_{i \in I_0(\Lambda^*)} (\Lambda_{\mathcal{M},i} - \Lambda_{\mathcal{M},*,i}) = \sum_{i \in I_0(\Lambda^*)} \Lambda_{\mathcal{M},i} = 0, \Lambda_{\mathcal{M},i} = \lambda_{\mathcal{M}}^T a_{\mathcal{M},i}. \end{aligned}$$

From (SM12.93) and the fact that  $\Lambda_{\mathcal{M}} \succeq 0$  it follows that

$$(SM12.94) \quad \Lambda_{\mathcal{M},i} = u_{\mathcal{M}}^T a_{\mathcal{M},i} = 0, i \in I_0(\Lambda^*).$$

Putting formulas (SM12.92) and (SM12.94) together, we arrive to the following:

$$(SM12.95) \quad u_{\mathcal{M}}^T a_{\mathcal{M},i} = 0 \text{ for } i \in \{1, \dots, d\}.$$

The injectivity of  $A_{\mathcal{M}}$  and (SM12.95) imply that  $u_{\mathcal{M}} = 0$  which contradicts the initial assumption that  $\|u_{\mathcal{M}}\| = \delta > 0$ .

Lemma is proved. ■

*Proof of Lemma SM12.7.* We prove the claim by contradiction.

The claim is obvious for  $\lambda_{\mathcal{M},*} = 0$ .

Let  $\lambda_{\mathcal{M},*} \neq 0$  and

$$(SM12.96) \quad \delta_* = \frac{1}{2} \min\{\lambda_{\mathcal{M},*,j} \mid \lambda_{\mathcal{M},*,j} > 0\}, \delta_* > 0.$$

Let  $u_{\mathcal{M}}$  be such that

$$(SM12.97) \quad 0 < \|u_{\mathcal{M}}\| \leq \delta_*, \lambda_{\mathcal{M},*} + u_{\mathcal{M}} \succeq 0$$

and assume that

$$(SM12.98) \quad \lambda_{\mathcal{M},*} + \delta_* \frac{u_{\mathcal{M}}}{\|u_{\mathcal{M}}\|} \not\succeq 0 \Leftrightarrow \exists j \in \{1, \dots, p_{\mathcal{M}}\} \text{ such that } \lambda_{\mathcal{M},*,j} + \delta_* \frac{u_{\mathcal{M},j}}{\|u_{\mathcal{M}}\|} < 0.$$

From the fact that  $\lambda_{\mathcal{M},*} \succeq 0$  and (SM12.97) and (SM12.98) it follows that

$$(SM12.99) \quad \text{for } j \text{ from (SM12.98) it holds that } \lambda_{\mathcal{M},*,j} > 0, u_{\mathcal{M},j} < 0.$$

Using (SM12.96), (SM12.98), and (SM12.99) we get the following implication:

$$(SM12.100) \quad \frac{\delta_*}{\|u_{\mathcal{M}}\|} (-u_{\mathcal{M},j}) > \lambda_{\mathcal{M},*,j} \geq 2\delta_* \Rightarrow (-u_{\mathcal{M},j}) > 2\|u_{\mathcal{M}}\|.$$

The inequality in the right hand-side of (SM12.100) gives the desired contradiction.

Lemma is proved. ■

**SM12.9. Proof of Theorem 6.10.**

*Proof.* In what follows we use the following auxiliary result.

**Theorem SM12.8 (concentration rate for the mixing parameter).** *Let Assumptions 6.3 and 6.7 be satisfied. Let  $\tilde{\lambda}_{\mathcal{M}}^t$  be sampled as in Algorithm 3 and  $r(t) = o(\sqrt{t/\log \log t})$ . Then,*

$$(SM12.101) \quad r(t)(\tilde{\lambda}_{\mathcal{M}}^t - \lambda_{\mathcal{M},*}) \xrightarrow{c.p.} 0,$$

where  $\lambda_{\mathcal{M},*}$  is from Theorem 6.9. In particular, formula (SM12.101) also implies

$$(SM12.102) \quad r(t)(\tilde{\Lambda}_{\mathcal{M}}^t - \Lambda_{\mathcal{M}}^*) \xrightarrow{c.p.} 0,$$

where  $\tilde{\Lambda}_{\mathcal{M}}^t = A\tilde{\lambda}_{\mathcal{M}}^t$ ,  $\Lambda_{\mathcal{M}}^* = A_{\mathcal{M}}\lambda_{\mathcal{M},*}$ .

**Remark SM12.9.** The log-factor for  $r(t)$  in Theorem SM12.8 is necessary for the ‘‘almost sure’’ character of formula (SM12.102) and, in particular, it is due to the Law of the Iterated Logarithm for trajectory  $Y^t$  (see section SM2). For our purposes it is sufficient to have the result for rate  $r(t) = o(\sqrt{t/\log \log t})$  because  $\tilde{\Lambda}_{\mathcal{M}}^t$  is used in the prior whose effect asymptotically disappears in view of the well-known Bernstein von-Mises phenomenon for Bayesian posteriors; see, e.g. Section 10.2 in [SM29].

The formula for  $\tilde{\lambda}_b^t$  in Step 6 of Algorithm 4 can be rewritten as follows:

$$(SM12.103) \quad \tilde{\lambda}_b^t = \arg \min_{\lambda \geq 0} A^t(\lambda),$$

$$(SM12.104) \quad \begin{aligned} A^t(\lambda) &= L_p(\lambda \mid t\tilde{\Lambda}_b^t, A, t, \beta^t) - L_p(\hat{\lambda}_{sc}^t \mid t\hat{\Lambda}_{sc}^t, A, t, \beta^t) \\ &= \sum_{i \in I_1(\Lambda^*)} -t(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \log \left( \frac{\Lambda_i}{\hat{\Lambda}_{sc,i}^t} \right) \\ &\quad + \sum_{i \in I_1(\Lambda^*)} -t\hat{\Lambda}_{sc,i}^t \log \left( \frac{\Lambda_i}{\hat{\Lambda}_{sc,i}^t} \right) + t(\Lambda_i - \hat{\Lambda}_{sc,i}^t) \\ &\quad + \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t\Lambda_i) + t\Lambda_i \\ &\quad - \left( \sum_{i \in I_0(\Lambda^*)} -t\hat{\Lambda}_{sc,i}^t \log(t\hat{\Lambda}_{sc,i}^t) + t\hat{\Lambda}_{sc,i}^t \right) \\ &\quad + \beta^t(\varphi(\lambda) - \varphi(\hat{\lambda}_{sc}^t)), \quad \hat{\Lambda}_{sc}^t = A\hat{\lambda}_{sc}^t \end{aligned}$$

where  $\hat{\lambda}_{sc}^t$  is the strongly consistent estimator from (6.21)–(6.23).

To prove the claim, first, we approximate  $A^t(\lambda)$  with quadratic process  $B^t(\lambda)$  for which its minimizers have the same asymptotic distribution in the  $\text{Span}(A^T) \cap \mathbb{R}_+^p$  as for  $A^t(\lambda)$ . Second, using this approximation we establish the statements in (i), (ii), but for minimizers of  $B^t(\lambda)$  which together with the previous approximation argument completes the proof.



Approximations  $B^t(\lambda)$ ,  $\tilde{\lambda}_{b,app}^t$  of  $A^t(\lambda)$ ,  $\tilde{\lambda}_b^t$  are defined by the formulas:

$$(SM12.105) \quad B^t(\lambda) = \sum_{i \in I_1(\Lambda^*)} -t(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{\Lambda_i - \hat{\Lambda}_{sc,i}}{\hat{\Lambda}_{sc,i}} + t \frac{(\Lambda_i - \hat{\Lambda}_{sc,i})^2}{2\hat{\Lambda}_{sc,i}} \\ + \sum_{i \in I_0(\Lambda^*)} t\Lambda_i, \quad \Lambda_i = a_i^T \lambda.$$

$$(SM12.106) \quad \tilde{\lambda}_{b,app}^t = \arg \min_{\lambda \geq 0} B^t(\lambda).$$

Process  $B^t(\lambda)$  is flat in directions from  $\ker A$ , therefore, though  $\tilde{\lambda}_{b,app}^t$  in (SM12.106) always exists, it may not be unique, and, in general,  $\tilde{\lambda}_{b,app}^t$  is set-valued. In what follows, if not said otherwise, for  $\tilde{\lambda}_{b,app}^t$  one chooses any point from the set of minimizers (claims will automatically hold for all points in  $\tilde{\lambda}_{b,app}^t$ ).

It may happen that  $a_i^T \tilde{\lambda}_{b,app}^t = 0$  for some  $i \in I_0(\Lambda^*)$ , so  $A^t(\tilde{\lambda}_{b,app}^t)$ , in general, may not be defined due to the presence of logarithmic terms in (SM12.104). For this reason we approximate  $\tilde{\lambda}_{b,app}^t$  with another auxiliary point  $\tilde{\lambda}_{app}^t$  defined by the formula:

$$(SM12.107) \quad \tilde{\lambda}_{app}^t = \tilde{\lambda}_{b,app}^t + \sum_{i \in I_0(\Lambda^*)} \tilde{\Lambda}_{b,i}^t \frac{a_i}{\|a_i\|^2},$$

where  $\tilde{\Lambda}_b^t$  is from Step 5 of Algorithm 4. It is easy to check that value  $A^t(\tilde{\lambda}_{app}^t)$  is always well-defined (for  $x = 0$  we take convention that  $x \log x = 0$ ).

Let  $\mathcal{V}$ ,  $\mathcal{U}$  be the subspaces defined in (6.16) and (6.17), respectively. From (SM12.107) and the definition of  $\mathcal{V}$ ,  $\mathcal{U}$  it follows that

$$(SM12.108) \quad \Pi_{\mathcal{U}}(\tilde{\lambda}_{app}^t - \tilde{\lambda}_{b,app}^t) \equiv 0,$$

where  $\Pi_{\mathcal{U}}$  is defined in (6.17). For the approximation on  $\mathcal{V}$  the following result holds.

**Lemma SM12.10.** *Let  $\mathcal{V}$  be the subspace defined in (6.16),  $\Pi_{\mathcal{V}}$  be defined in (6.19). Then,*

$$(SM12.109) \quad t\Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \tilde{\lambda}_{app}^t) \xrightarrow{c.p.} 0.$$

Let  $\delta > 0$ . Consider the two following sets:

$$(SM12.110) \quad D_{A,\delta}^t(\lambda) = \{\lambda' \in \mathbb{R}_+^p : \lambda' = \lambda + \frac{u}{\sqrt{t}} + \frac{v}{t} + w, u \in \mathcal{U}, v \in \mathcal{V}, w \in \mathcal{W}, \|u\|_2 + \|v\|_1 \leq \delta\},$$

$$(SM12.111) \quad C_{A,\delta}^t(\lambda) = \{\lambda' \in \mathbb{R}_+^p : \lambda' = \lambda + \frac{u}{\sqrt{t}} + \frac{v}{t} + w, u \in \mathcal{U}, v \in \mathcal{V}, w \in \mathcal{W}, \|u\|_2 + \|v\|_1 = \delta\},$$

where subspaces  $\mathcal{V}, \mathcal{U}, \mathcal{W}$  are defined in (6.16)–(6.18), respectively and  $\|\cdot\|_2, \|\cdot\|_1$  denote the standard  $\ell_2$  and  $\ell_1$ -norms in  $\mathbb{R}^p$ .

The approximation argument for convex process  $A^t(\lambda)$  is due to [SM16] and is based on the following implication:

$$(SM12.112) \quad \tilde{\lambda}_{app}^t \in \text{int}D_{A,\delta}^t(\tilde{\lambda}_{b,app}^t), \quad \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} (A^t(\lambda) - A^t(\tilde{\lambda}_{app}^t)) > 0 \Rightarrow \tilde{\lambda}_b^t \in D_{A,\delta}^t(\tilde{\lambda}_{b,app}^t).$$

From (SM12.108) and (SM12.109) (in Lemma SM12.10) and (SM12.110) one can see that for any  $\delta > 0$  it holds that

$$(SM12.113) \quad P(\tilde{\lambda}_{app}^t \in \text{int}D_{A,\delta}^t(\tilde{\lambda}_{b,app}^t) \mid Y^t, t) \rightarrow 1 \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty).$$

In view of this and (SM12.112), for the approximation it suffices to establish the following result.

**Lemma SM12.11.** *Let  $A^t(\lambda)$ ,  $B^t(\lambda)$ ,  $\tilde{\lambda}_b^t$ ,  $\tilde{\lambda}_{b,app}^t$ ,  $\tilde{\lambda}_{app}^t$  be defined in (SM12.103), (SM12.104), (SM12.105), (SM12.106), and (SM12.107), respectively. Then, for any  $\delta > 0$  it holds that*

$$(SM12.114) \quad P\left(\inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - A^t(\tilde{\lambda}_{app}^t)] > 0 \mid Y^t, t\right) \rightarrow 1 \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty).$$

From (SM12.112) and (SM12.114) it follows that

$$(SM12.115) \quad \sqrt{t}\Pi_{\mathcal{U}}(\tilde{\lambda}_b^t - \tilde{\lambda}_{b,app}^t) \xrightarrow{c.p.} 0,$$

$$(SM12.116) \quad t\Pi_{\mathcal{V}}(\tilde{\lambda}_b^t - \tilde{\lambda}_{b,app}^t) \xrightarrow{c.p.} 0.$$

Let

$$(SM12.117) \quad \lambda = \tilde{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + \frac{v}{t} + w, \quad u \in \mathcal{U}, v \in \mathcal{V}, w \in \mathcal{W}.$$

Process  $B^t(\cdot)$  defined in (SM12.105) has the following form in terms of variables  $u, v$  (note that  $B^t(\cdot)$  is independent of  $w \in \mathcal{W}$ ):

$$(SM12.118) \quad B^t(u, v) = \tilde{B}^t(u, v) + \tilde{R}^t(u, v),$$

$$(SM12.119) \quad \tilde{B}^t(u, v) = \sum_{i \in I_1(\Lambda^*)} -\sqrt{t}(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{a_i^T u}{\hat{\Lambda}_{sc,i}^t} + \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_0(\Lambda^*)} a_i^T v,$$

$$(SM12.120) \quad \tilde{R}^t(u, v) = \sum_{i \in I_1(\Lambda^*)} -(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) a_i^T v + \frac{(a_i^T v)^2}{2\hat{\Lambda}_{sc,i}^t t} + \frac{(a_i^T u)(a_i^T v)}{\sqrt{t}\hat{\Lambda}_{sc,i}^t} \\ + \sum_{i \in I_0(\Lambda^*)} t\hat{\Lambda}_{sc,i}^t.$$

Let

$$(SM12.121) \quad (\tilde{u}^t, \tilde{v}^t) = \arg \min_{\substack{(u,v): \tilde{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + \frac{v}{t} + w \geq 0 \\ u \in \mathcal{U}, v \in \mathcal{V}, w \in \mathcal{W}}} \tilde{B}^t(u, v)$$

In particular, from the definition of  $\mathcal{V}$  in (6.16) and from (SM12.117), (SM12.119), and (SM12.121) it follows that

$$(SM12.122) \quad \frac{\tilde{v}_j^t}{t} = -\hat{\lambda}_{sc,j}^t \text{ for } j \text{ s.t. } \exists a_{ij} > 0, i \in I_0(\Lambda^*) \Leftrightarrow \Pi_{\mathcal{V}}(\hat{\lambda}_{sc}^t + \frac{\tilde{v}^t}{t}) = 0.$$

Indeed, formulas (6.16), (6.19), and (SM12.119) imply that the choice in (SM12.122) satisfies the positivity constraint in (SM12.121) and at the same time minimizes the linear term  $\sum_{i \in I_0(\Lambda^*)} a_i^T v$  since all  $a_{ij}$  are non-negative.

**Lemma SM12.12.** *Let  $\tilde{u}_{b,app}^t, \tilde{v}_{b,app}^t$  be defined by (SM12.106) for parametrization in formula (SM12.117) and  $\tilde{u}^t, \tilde{v}^t$  be defined by (SM12.121), respectively. Then,*

$$(SM12.123) \quad \tilde{u}^t - \tilde{u}_{b,app}^t \xrightarrow{c.p.} 0 \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty),$$

$$(SM12.124) \quad \tilde{v}^t - \tilde{v}_{b,app}^t \xrightarrow{c.p.} 0 \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty).$$

Hence, in view of (SM12.115), (SM12.116) and Lemma SM12.12 it suffices to demonstrate conditional tightness of  $(\tilde{u}^t, \tilde{v}^t)$ .

Statement in (i), that is formula (6.24), follows from (SM12.116), (SM12.122), and (SM12.124) and the assumption in (6.23).

Now we demonstrate (ii). From (SM12.121), (SM12.122) it follows that

$$(SM12.125) \quad \tilde{u}^t = \arg \min_{\substack{u: (1-\Pi_{\mathcal{V}})\hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + w \succeq 0 \\ u \in \mathcal{U}, w \in \mathcal{W}}} \sum_{i \in I_1(\Lambda^*)} -\sqrt{t}(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{a_i^T u}{\hat{\Lambda}_{sc,i}^t} + \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t}.$$

Since the minimized functional in (SM12.125) is strongly convex in  $u \in \mathcal{U}$  and the set of constraints is also convex, the following mapping is well-defined:

$$(SM12.126) \quad \tilde{u}^t(\xi) = \tilde{u}(\xi, t) \in \mathcal{U}, \xi \in \mathbb{R}^{\#I_1(\Lambda^*)}, t \in (0, +\infty),$$

$$(SM12.127) \quad \tilde{u}(\xi, t) = \arg \min_{\substack{u: (1-\Pi_{\mathcal{V}})\hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + w \succeq 0 \\ u \in \mathcal{U}, w \in \mathcal{W}}} -\xi^T (\hat{D}_{I_1(\Lambda^*)}^t)^{-1/2} A_{I_1(\Lambda^*)} u + \frac{1}{2} u^T \hat{F}_{I_1(\Lambda^*)}^t u,$$

where

$$(SM12.128) \quad \hat{D}_{I_1(\Lambda^*)}^t = \text{diag}(\dots, \hat{\Lambda}_{sc,i}^t, \dots), i \in I_1(\Lambda^*),$$

$$(SM12.129) \quad \hat{F}_{I_1(\Lambda^*)}^t = \sum_{i \in I_1(\Lambda^*)} \frac{a_i a_i^T}{\hat{\Lambda}_{sc,i}^t} = A_{I_1(\Lambda^*)}^T (\hat{D}_{I_1(\Lambda^*)}^t)^{-1} A_{I_1(\Lambda^*)}.$$

Note that for  $\xi = (\dots, \sqrt{t}(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t)/\sqrt{\hat{\Lambda}_{sc,i}^t}, \dots)$ ,  $i \in I_1(\Lambda^*)$ ,  $\tilde{u}^t(\xi)$  coincides with  $\tilde{u}^t$  from (SM12.125). In addition, the minimized functional in (SM12.127) does not depend on  $w \in \mathcal{W}$  which in turn affects only the set of constraints.

**Lemma SM12.13.** *Let  $\tilde{u}^t(\xi)$  be the mapping defined in (SM12.126)–(SM12.129). Then,*

$$(SM12.130) \quad \|\tilde{u}^t(\xi)\| \leq \tilde{c}^t \|A_{I_1(\Lambda^*)}^T \widehat{D}_{I_1(\Lambda^*)}^t \xi\|, \quad \xi \in \mathbb{R}^{\#I_1(\Lambda^*)},$$

$$(SM12.131) \quad \tilde{c}^t = \|(\widehat{F}_{I_1(\Lambda^*)}^t)^{-1}\|_{\mathcal{U}} \cdot \|(\widehat{F}_{I_1(\Lambda^*)}^t)^{-1/2}\| \left( \|(\widehat{F}_{I_1(\Lambda^*)}^t)^{-1}\|_{\mathcal{U}} + 2 \max_{\sigma \in \sigma_{\mathcal{U}}(\widehat{F}_{I_1(\Lambda^*)}^t)} \sigma^{-1/2} \right),$$

where  $\|\cdot\|_{\mathcal{U}}$  denotes the norm of the operator being reduced to subspace  $\mathcal{U}$ ,  $\sigma_{\mathcal{U}}(\cdot)$  denotes the spectrum of the self-adjoint operator acting on  $\mathcal{U}$ . Moreover,

$$(SM12.132) \quad \tilde{c}^t \rightarrow c^* \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty), c^* < +\infty,$$

$$(SM12.133) \quad c_* = \|(F_{I_1(\Lambda^*)}^*)^{-1}\|_{\mathcal{U}} \cdot \|(F_{I_1(\Lambda^*)}^*)^{-1/2}\| \left( \|(F_{I_1(\Lambda^*)}^*)^{-1}\|_{\mathcal{U}} + 2 \max_{\sigma \in \sigma_{\mathcal{U}}(F_{I_1(\Lambda^*)}^*)} \sigma^{-1/2} \right),$$

where

$$(SM12.134) \quad D_{I_1(\Lambda^*)} = \text{diag}(\dots, \Lambda_i^*, \dots), \quad i \in I_1(\Lambda^*),$$

$$(SM12.135) \quad F_{I_1(\Lambda^*)}^* = \sum_{i \in I_1(\Lambda^*)} \frac{a_i a_i^T}{\Lambda_i^*} = A_{I_1(\Lambda^*)}^T D_{I_1(\Lambda^*)}^{-1} A_{I_1(\Lambda^*)}.$$

**Lemma SM12.14.** *Let*

$$(SM12.136) \quad \tilde{\xi}^t = (\dots, \sqrt{t}(\tilde{\Lambda}_{b,i}^t - \widehat{\Lambda}_{sc,i}^t) / \sqrt{\widehat{\Lambda}_{sc,i}^t}, \dots), \quad i \in I_1(\Lambda^*), \tilde{\xi}^t \in \mathbb{R}^{\#I_1(\Lambda^*)}.$$

Then, under the assumptions of [Theorem 6.10](#), family  $A_{I_1(\Lambda^*)}^T (\widehat{D}_{I_1(\Lambda^*)}^t)^{-1/2} \tilde{\xi}^t$  is conditionally tight.

The result of [Lemma SM12.14](#) together with formulas (SM12.130)–(SM12.135) imply that  $\tilde{u}^t = \tilde{u}^t(\tilde{\xi}^t)$  is conditionally tight almost surely  $Y^t$ ,  $t \in (0, +\infty)$ . Statement (ii) of the lemma follows directly from this and formulas (SM12.115), (SM12.123) from [Lemmas SM12.11](#) and [SM12.12](#), respectively.

Theorem is proved.

### SM12.10. Proof of [Theorem SM12.8](#).

*Proof.* Claim in (SM12.102) directly follows from (SM12.101) and the Continuous Mapping Theorem, so we prove only (SM12.101).

Step 4 in [Algorithm 3](#) can be rewritten as follows:

$$(SM12.137) \quad \tilde{\lambda}_{\mathcal{M}}^t = \arg \min_{\lambda_{\mathcal{M}} \geq 0} L_{\mathcal{M}}(\lambda_{\mathcal{M}} \mid \tilde{\Lambda}^t),$$

$$(SM12.138) \quad L_{\mathcal{M}}(\lambda_{\mathcal{M}} \mid \tilde{\Lambda}^t) = \sum_{i \in I_1(\Lambda^*)} -\log \left( \frac{\Lambda_{\mathcal{M},i}}{\Lambda_{\mathcal{M},i}^*} \right) (\tilde{\Lambda}_i^t - \Lambda_i^*) \\ + L(\lambda_{\mathcal{M}} \mid \Lambda^*, A_{\mathcal{M}}, 1) - L(\lambda_{\mathcal{M},*} \mid \Lambda^*, A_{\mathcal{M}}, 1),$$

where  $\lambda_{\mathcal{M},*}$  is the point from [Theorem 6.9](#),  $\Lambda_{\mathcal{M}}^* = A_{\mathcal{M}}\lambda_{\mathcal{M},*}$ , and

$$(SM12.139) \quad \begin{aligned} \tilde{\Lambda}_i^t &\sim \Gamma(Y_i^t, t^{-1}), \quad i = 1, \dots, d, \text{ are mutually independent,} \\ E[\tilde{\Lambda}_i^t | Y^t, t] &= Y_i^t/t, \quad \text{var}[\tilde{\Lambda}_i^t | Y^t, t] = Y_i^t/t^2, \quad i \in \{1, \dots, d\}. \end{aligned}$$

Note that

$$(SM12.140) \quad L_{\mathcal{M}}(\lambda_{\mathcal{M}} | \tilde{\Lambda}^t) \text{ is convex on } \mathbb{R}_+^{p_{\mathcal{M}}}, \quad L_{\mathcal{M}}(\lambda_{\mathcal{M},*} | \tilde{\Lambda}^t) = 0.$$

For fixed  $t > 0$  consider the following parametrization

$$(SM12.141) \quad \lambda_{\mathcal{M}} = \lambda_{\mathcal{M},*} + \frac{u_{\mathcal{M}}}{r(t)}, \quad \lambda_{\mathcal{M}} \in \mathbb{R}_+^{p_{\mathcal{M}}}, \quad r(t) = o(\sqrt{t/\log \log t}).$$

Let  $\delta > 0$ . In view of [\(SM12.137\)](#), [\(SM12.140\)](#), and [\(SM12.141\)](#) the following implication holds

$$(SM12.142) \quad \inf_{\substack{\lambda_{\mathcal{M}}: \|u_{\mathcal{M}}\|=\delta, \\ \lambda_{\mathcal{M}} \geq 0}} L_{\mathcal{M}}(\lambda_{\mathcal{M}} | \tilde{\Lambda}^t) > 0 \Rightarrow r(t) \|\tilde{\lambda}_{\mathcal{M}}^t - \lambda_{\mathcal{M},*}\| < \delta.$$

Therefore, to prove [\(SM12.101\)](#) it is sufficient to show that for any small  $\delta > 0$  the conditional probability of the event in the left hand-side of [\(SM12.142\)](#) tends to one for  $t \rightarrow +\infty$ , a.s.  $Y^t, t \in [0, +\infty)$ .

Let  $C_*, \delta_*$  be the values of [\(6.15\)](#) from [Theorem 6.9](#) and let  $\|u_{\mathcal{M}}\| = \delta, \delta < \delta_*$ .

Using [\(6.15\)](#) and [\(SM12.138\)](#), [\(SM12.141\)](#) we get the following estimate:

$$(SM12.143) \quad \begin{aligned} L(\lambda_{\mathcal{M}} | \tilde{\Lambda}^t) &\geq \sum_{i \in I_1(\Lambda^*)} -\log \left( 1 + \frac{u_{\mathcal{M}}^T a_{\mathcal{M},i}}{r(t)\Lambda_{\mathcal{M},i}^*} \right) (\tilde{\Lambda}_i^t - \Lambda_i^*) + C_*\delta^2/r^2(t) \\ &\geq C_*\delta^2/r^2(t) - \sum_{i \in I_1(\Lambda^*)} \frac{|u_{\mathcal{M}}^T a_{\mathcal{M},i}|}{r(t)\Lambda_{\mathcal{M},i}^*} |\tilde{\Lambda}_i^t - \Lambda_i^*| \\ &= r^{-2}(t) \left( C_*\delta^2 - \sum_{i \in I_1(\Lambda^*)} \frac{|u_{\mathcal{M}}^T a_{\mathcal{M},i}|}{\Lambda_{\mathcal{M},i}^*} r(t) |\tilde{\Lambda}_i^t - \Lambda_i^*| \right) \\ &\geq r^{-2}(t) \left( C_*\delta^2 - \sum_{i \in I_1(\Lambda^*)} \frac{\delta \|a_{\mathcal{M},i}\|}{\Lambda_{\mathcal{M},i}^*} r(t) |\tilde{\Lambda}_i^t - \Lambda_i^*| \right). \end{aligned}$$

Note that in [\(SM12.143\)](#) we have used the property that  $\log(1+x) \leq x, x \in (-1, +\infty)$ .

Estimate in [\(SM12.143\)](#) implies the left hand-side of [\(SM12.142\)](#), for example, if

$$(SM12.144) \quad r(t) |\tilde{\Lambda}_i^t - \Lambda_i^*| \xrightarrow{c.p.} 0, \quad i \in I_1(\Lambda^*).$$

To demonstrate (SM12.144) we use Markov inequality together with (SM12.139) and arrive to the following estimate

(SM12.145)

$$\begin{aligned} P(r(t) | \tilde{\Lambda}_i^t - \Lambda_i^* | > \varepsilon | Y^t, t) &\leq \frac{r^2(t) E(|\tilde{\Lambda}_i^t - \Lambda_i^*|^2 | Y^t, t)}{\varepsilon^2} \\ &\leq \frac{2r^2(t) E(|\tilde{\Lambda}_i^t - Y_i^t/t|^2 | Y^t, t) + 2r^2(t) |Y_i^t - \Lambda_i^*|^2}{\varepsilon^2} \\ &= \frac{2r^2(t)/t^2 + 2 |r(t)(Y_i^t/t - \Lambda_i^*)|^2}{\varepsilon^2}, \end{aligned}$$

where  $\varepsilon > 0$  is arbitrary. For  $r(t) = o(\sqrt{t/\log \log t})$  it holds that (see section SM2):

$$(SM12.146) \quad r^2(t)/t^2 \rightarrow 0 \text{ and } r(t)(Y_i^t/t - \Lambda_i^*) \rightarrow 0 \text{ a.s. } Y^t, t \in (0, +\infty). \quad \blacksquare$$

Therefore, from (SM12.145), (SM12.146) it follows that formula (SM12.144) holds which together with (SM12.143) imply (SM12.142).

Theorem is proved.

### SM12.11. Proof of Lemma SM12.10.

*Proof.* To prove the claim it suffices to show that

$$(SM12.147) \quad t\tilde{\Lambda}_{b,i}^t \xrightarrow{c.p.} 0 \text{ for } i \in I_0(\Lambda^*).$$

Let  $\delta > 0$ . Using Step 5 of Algorithm 4 and Assumption 6.3 we obtain

$$\begin{aligned} P(t\tilde{\Lambda}_{b,i}^t > \delta | Y^t, t) &= \int_0^{+\infty} P(t\tilde{\Lambda}_{b,i}^t > \delta | \tilde{\Lambda}_{\mathcal{M},i}^t = \Lambda, Y^t, t) P(\tilde{\Lambda}_{\mathcal{M},i}^t = \Lambda | Y^t, t) d\Lambda \\ (SM12.148) \quad &\leq \int_0^{+\infty} \min\left(\frac{t\theta^t \Lambda}{(\theta^t + t)\delta}, 1\right) P(\tilde{\Lambda}_{\mathcal{M},i}^t = \Lambda | Y^t, t) d\Lambda \\ &\leq \int_0^{\frac{(\theta^t + t)\delta}{t\theta^t}} \frac{t\theta^t \Lambda}{(\theta^t + t)\delta} P(\tilde{\Lambda}_{\mathcal{M},i}^t = \Lambda | Y^t, t) d\Lambda + P\left(\frac{t\theta^t \tilde{\Lambda}_{\mathcal{M},i}^t}{\theta^t + t} > \delta | Y^t, t\right). \end{aligned}$$

In (SM12.148) we have used the Markov inequality for  $\Lambda_b^t | Y^t, t, \tilde{\Lambda}_{\mathcal{M},i}^t, i \in I_0(\Lambda^*)$  for which it is known that  $\tilde{\Lambda}_{b,i}^t | Y^t, t, \tilde{\Lambda}_{\mathcal{M},i}^t \sim \Gamma(\theta^t \tilde{\Lambda}_{\mathcal{M},i}^t, (t + \theta^t)^{-1})$ .

The last term in (SM12.148) tends to zero a.s.  $Y^t, t \in (0, +\infty)$  due to (SM12.102) from Theorem SM12.8.

Next, we show that the first integral in (SM12.148) it is arbitrarily small a.s.  $Y^t, t \in [0, +\infty)$  and, hence, tends to zero a.s.  $Y^t, t \in (0, +\infty)$ . The integral in (SM12.148) is

rewritten as follows:

$$\begin{aligned}
 (\text{SM12.149}) \quad & \int_0^{\frac{(\theta^t+t)\delta}{t\theta^t}} \frac{t\theta^t\Lambda}{(\theta^t+t)\delta} P(\tilde{\Lambda}_{\mathcal{M},i}^t = \Lambda \mid Y^t, t) d\Lambda = \\
 & = \frac{\delta(\theta^t+t)}{t\theta^t} \int_0^1 s P(\theta^t \tilde{\Lambda}_{\mathcal{M},i}^t = s\delta(t+\theta^t)/t \mid Y^t, t) ds.
 \end{aligned}$$

Let  $0 < \varepsilon < 1$ . Then, by splitting the integral in (SM12.149) we obtain the following estimate:

$$\begin{aligned}
 (\text{SM12.150}) \quad & \frac{\delta(\theta^t+t)}{t\theta^t} \int_0^1 s P(\theta^t \Lambda_{\mathcal{M},i}^t = s\delta(t+\theta^t)/t \mid Y^t, t) ds = \int_0^\varepsilon \dots ds + \int_\varepsilon^1 \dots ds \\
 & \leq \varepsilon + P(\theta^t \Lambda_{\mathcal{M},i}^t > \varepsilon\delta(t+\theta^t)/t \mid Y^t, t).
 \end{aligned}$$

For fixed  $\varepsilon > 0$ ,  $\delta > 0$ , the second term in (SM12.150) tends to zero for  $t \rightarrow +\infty$ , a.s.  $Y^t$ ,  $t \in (20, +\infty)$ , again due to (SM12.102) from Theorem SM12.8. Since  $\varepsilon$  can be arbitrarily small, it follows that the integral in (SM12.150) is also arbitrarily small for  $t \rightarrow +\infty$ , a.s.  $Y^t$ ,  $t \in (0, +\infty)$ . Hence, the integral in (SM12.149), and most importantly the right hand-side in (SM12.148) converge to zero when  $t \rightarrow +\infty$ , a.s.  $Y^t$ ,  $t \in (0, +\infty)$ . Since initial  $\delta$  was chosen arbitrarily, this proves the convergence in (SM12.147).

Lemma is proved. ■

### SM12.12. Proof of Lemma SM12.11.

*Proof.* Let  $\delta > 0$ . The left hand-side of (SM12.112) can be estimated as follows:

$$\begin{aligned}
 (\text{SM12.151}) \quad & \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - A^t(\tilde{\lambda}_{app}^t)] \geq \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - B^t(\lambda)] \\
 & \quad + \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [B^t(\lambda) - B^t(\tilde{\lambda}_{b,app}^t)] \\
 & \quad + [B^t(\tilde{\lambda}_{b,app}^t) - B^t(\tilde{\lambda}_{app}^t)] \\
 & \quad + [B^t(\tilde{\lambda}_{app}^t) - A^t(\tilde{\lambda}_{app}^t)].
 \end{aligned}$$

We will show that under the assumptions of Theorem 6.10 the following holds:

$$(\text{SM12.152}) \quad \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - A^t(\tilde{\lambda}_{app}^t)] \geq \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [B^t(\lambda) - B^t(\tilde{\lambda}_{b,app}^t)] + o_{cp}(1).$$

The first term in right hand-side of (SM12.152) is expected to be positively separated from zero in view of (SM12.106), (SM12.111), and in fact, it gives the main contribution for (SM12.112) to hold. This is described precisely by the following lemma.

**Lemma SM12.15.** *Let  $B^t(\lambda)$ ,  $\tilde{\lambda}_{b,app}^t$  be defined in (SM12.105) and (SM12.106), respectively. Then, the following formulas hold:*

$$(SM12.153) \quad B^t(\lambda) - B^t(\tilde{\lambda}_{b,app}^t) = \sum_{i \in I_1(\Lambda^*)} \frac{t(\Lambda_i - \tilde{\Lambda}_{b,app,i}^t)^2}{2\hat{\Lambda}_{sc,i}^t} + t\langle \tilde{\mu}_{b,app}^t, \lambda \rangle,$$

$$\lambda \in \mathbb{R}_+^p, \tilde{\Lambda}_{b,app}^t = A\tilde{\lambda}_{b,app}^t,$$

where

$$(SM12.154) \quad \tilde{\mu}_{b,app}^t = \sum_{i \in I_1(\Lambda^*)} \frac{\tilde{\Lambda}_{b,app,i}^t - \tilde{\Lambda}_{b,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i + \sum_{i \in I_0(\Lambda^*)} a_i,$$

$$(SM12.155) \quad \tilde{\mu}_{b,app}^t \in \mathbb{R}_+^p, \tilde{\mu}_{b,app,j}^t \tilde{\lambda}_{b,app,j}^t = 0 \text{ for all } j \in \{1, \dots, p\}.$$

We show that (SM12.152) and the result of Lemma SM12.15 imply the statement in formula (SM12.114).

Let

$$(SM12.156) \quad \lambda(u, v, w) = \tilde{\lambda}_{b,app}^t + \frac{u}{\sqrt{t}} + \frac{v}{t} + w, \quad u \in \mathcal{U}, v \in \mathcal{V}, w \in \mathcal{W}, \lambda(u, v, w) \in \mathbb{R}_+^p.$$

Using the parametrization from (SM12.156), the definition of  $C_{A,\delta}^t(\cdot)$  in (SM12.111) and (SM12.153)–(SM12.155) from (SM12.152) we obtain

$$(SM12.157) \quad B^t(\lambda) - B^t(\tilde{\lambda}_{b,app}^t) = K^t(u, v, w) + R^t(u, v, w), \quad \lambda = \lambda(u, v, w),$$

$$(SM12.158) \quad \begin{aligned} K^t(u, v, w) &= \sum_{i \in I_1(\Lambda^*)} \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t} + t\langle \tilde{\mu}_{b,app}^t, \lambda(u, v, w) \rangle \\ &= \sum_{i \in I_1(\Lambda^*)} \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t} + t\langle \tilde{\mu}_{b,app}^t, \lambda(u, v, w) - \tilde{\lambda}_{b,app}^t \rangle \\ &= \sum_{i \in I_1(\Lambda^*)} \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t} + t\langle \tilde{\mu}_{b,app}^t, \frac{u}{\sqrt{t}} + \frac{v}{t} \rangle \\ &= \sum_{i \in I_1(\Lambda^*)} \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_1(\Lambda^*)} \sqrt{t} \frac{\tilde{\Lambda}_{b,app,i}^t - \tilde{\Lambda}_{b,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i^T u \\ &\quad + \sum_{i \in I_1(\Lambda^*)} \frac{\tilde{\Lambda}_{b,app,i}^t - \tilde{\Lambda}_{b,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i^T v + \sum_{i \in I_0(\Lambda^*)} a_i^T v, \end{aligned}$$

$$(SM12.159) \quad R^t(u, v, w) = \sum_{i \in I_1(\Lambda^*)} \frac{(a_i^T u)(a_i^T v)}{\sqrt{t}\hat{\Lambda}_{sc,i}^t} + \frac{(a_i^T v)^2}{2t\hat{\Lambda}_{sc,i}^t}.$$

From the fact that  $\hat{\Lambda}_{sc,i}^t \rightarrow \Lambda_i^*$  a.s.  $Y^t, t \in [0, +\infty)$  ( $\hat{\Lambda}_{sc}^t$  is strongly consistent at  $\lambda_*$  on  $\mathcal{U} \oplus \mathcal{V}$  by the assumption), the definition of  $C_{A,\delta}^t(\cdot)$  in (SM12.111) and (SM12.159) it follows that

$$(SM12.160) \quad \sup_{\lambda(u,v,w) \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} |R^t(u, v, w)| = o_{cp}(1).$$



In view of formulas (SM12.117) and (SM12.122), the results of Lemmas SM12.12 to SM12.14 and again the fact that  $\widehat{\Lambda}_{sc,i}^t \rightarrow \Lambda_i^*$ , we find that

$$(SM12.161) \quad \frac{\widetilde{\Lambda}_{b,app,i}^t - \widetilde{\Lambda}_{b,i}^t}{\widehat{\Lambda}_{sc,i}^t} = \frac{\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t}{\widehat{\Lambda}_{sc,i}^t} + \frac{\widehat{\Lambda}_{sc,i}^t - \widetilde{\Lambda}_{b,i}^t}{\widehat{\Lambda}_{sc,i}^t} = o_{cp}(1), \quad i \in I_1(\Lambda^*).$$

Formulas (SM12.157)–(SM12.161) imply that

$$(SM12.162) \quad \inf_{\lambda \in C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t)} [B^t(\lambda) - B^t(\widetilde{\lambda}_{b,app}^t)] \geq \inf_{\lambda(u,v,w) \in C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t)} K^t(u,v,w) + o_{cp}(1).$$

Now, note that if  $\lambda(u,v,w) \succeq 0$  (see formula (SM12.156)), then

$$(SM12.163) \quad \lambda(u, 0, w) \succeq 0.$$

Indeed, from the definition of  $\mathcal{V}, \mathcal{U}, \mathcal{W}$  in (6.16)–(6.18) it follows that  $u$  and  $v$  have disjoint set of non-zero components, therefore, setting  $v$  to zero for  $\lambda(u,v,w)$  cannot break the positivity constraint.

From (SM12.154), (SM12.155), and (SM12.163) it follows that

$$(SM12.164) \quad \langle \widetilde{\mu}_{b,app}^t, \lambda(u, 0, w) \rangle = \sum_{i \in I_1(\Lambda^*)} \frac{\widetilde{\Lambda}_{b,app,i}^t - \widetilde{\Lambda}_{b,i}^t}{\widehat{\Lambda}_{sc,i}^t} a_i^T u \geq 0.$$

Note also that  $K^t(u,v,w)$  in (SM12.158) does not change when varying  $w \in \mathcal{W}$ , so, in what follows we write  $K^t(u,v)$  instead. Using formulas (SM12.158), (SM12.161), (SM12.164) and the definition of  $C_{A,\delta}^t(\cdot)$  in (SM12.111) we find that

$$(SM12.165) \quad K^t(u,v) \geq \sum_{i \in I_1(\Lambda^*)} \frac{(a_i^T u)^2}{2\widehat{\Lambda}_{sc,i}^t} + \sum_{i \in I_0(\Lambda^*)} a_i^T v + o_{cp}(1), \quad \lambda = \lambda(u,v,w) \in C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t).$$

where the term  $o_{cp}(1)$  tends to zero uniformly on  $C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t)$  for  $t \rightarrow +\infty$ , a.s.  $Y^t$ ,  $t \in (0, +\infty)$ . From (SM12.165) and strong consistency of  $\widehat{\lambda}_{sc}^t$  on  $\mathcal{U} \oplus \mathcal{V}$  it follows that

$$(SM12.166) \quad K^t(u,v) \geq c_1 \|u\|_2^2 + c_2 \|v\|_1 + o_{cp}(1), \quad \text{if } \Pi_{\mathcal{V}} \widetilde{\lambda}_{b,app}^t = 0, \quad \lambda = \lambda(u,v,w) \in C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t),$$

where  $c_1, c_2$  are some fixed positive constants which depend only on  $\Lambda^*$  and  $A$ . The bound above holds for  $t$  large enough a.s.  $Y^t$ ,  $t \in (0, +\infty)$ .

Recall that

$$(SM12.167) \quad \|u\|_2 + \|v\|_1 = \delta \quad \text{for } \lambda(u,v,w) \in C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t).$$

Using (SM12.166), (SM12.167) it is easy to see that

$$(SM12.168) \quad K^t(u,v) \geq c\delta^2 + o_{cp}(1), \quad \text{if } \Pi_{\mathcal{V}} \widetilde{\lambda}_{b,app}^t = 0, \quad \lambda = \lambda(u,v,w) \in C_{A,\delta}^t(\widetilde{\lambda}_{b,app}^t),$$

for  $\delta$  small enough (smaller than some universal constant depending on  $c_1, c_2$ ), where  $c$  is some fixed constant also depending on  $c_1, c_2$  from (SM12.166). Note that the Karush-Kuhn-Tucker optimality conditions in (SM12.154), (SM12.155), formula (SM12.161) and the definition of space  $\mathcal{V}$  in (6.16) imply that

$$(SM12.169) \quad P(\Pi_{\mathcal{V}} \tilde{\lambda}_{b,app}^t = 0 \mid Y^t, t) \rightarrow 1 \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty).$$

Hence, the event in (SM12.168) is conditioned on  $\{\Pi_{\mathcal{V}} \tilde{\lambda}_{b,app}^t = 0\}$  which has asymptotic conditional probability tending to one a.s.  $Y^t, t \in [0, +\infty)$ , and it also holds

$$(SM12.170) \quad K^t(u, v) \geq c\delta^2 + o_{cp}(1), \lambda = \lambda(u, v, w) \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t).$$

From (SM12.152), (SM12.162), and (SM12.170) it follows that

$$(SM12.171) \quad P\left(\inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - A^t(\tilde{\lambda}_{b,app}^t)] > 0 \mid Y^t, t\right) \rightarrow 1 \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty).$$

It is left to demonstrate the initial statement in (SM12.152). Consider the first term in the left hand-side of (SM12.151). Using (SM12.104), (SM12.105), the definitions in (SM12.106), (SM12.111) and the facts that  $\Pi_{\mathcal{V} \oplus \mathcal{U}}(\hat{\lambda}_{sc}^t - \lambda_*) \xrightarrow{a.s.} 0$ ,  $\Pi_{\mathcal{V} \oplus \mathcal{U}}(\tilde{\lambda}_{b,app}^t - \lambda_*) \xrightarrow{c.p.} 0$ , and the Taylor expansion of  $A(\lambda)$  at  $\hat{\lambda}_{sc}^t$  up to the second order one gets the following estimate

$$(SM12.172) \quad \begin{aligned} A^t(\lambda) - B^t(\lambda) &\geq \sum_{i \in I_1(\Lambda^*)} -tC_1 |\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t| \frac{|\Lambda_i - \hat{\Lambda}_{sc,i}^t|^2}{|\hat{\Lambda}_{sc,i}^t|^2} + \sum_{i \in I_1(\Lambda^*)} -tC_2 |\Lambda_i - \hat{\Lambda}_{sc,i}^t|^3 \\ &+ \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t\Lambda_i) + \sum_{i \in I_0(\Lambda^*)} t\hat{\Lambda}_{sc,i}^t \log(t\hat{\Lambda}_{sc,i}^t) - t\hat{\Lambda}_{sc,i}^t \\ &+ \beta^t(\varphi(\lambda) - \varphi(\hat{\lambda}_{sc}^t)), \lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t). \end{aligned}$$

where  $C_1, C_2$  are some positive constants which depend only design  $A$  and  $\Lambda^*$ . The above estimate holds with conditional probability tending to one for  $t \rightarrow +\infty$  a.s.  $Y^t, t \in [0, +\infty)$ . In particular, in (SM12.172) to bound uniformly the error-terms in the Taylor's expansion we have used the following estimates:

$$(SM12.173) \quad \sup_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} |\Lambda_i - \hat{\Lambda}_{sc,i}^t| / |\hat{\Lambda}_{sc,i}^t| = o_{cp}(1), i \in I_1(\Lambda^*),$$

$$(SM12.174) \quad |\log(1+x) - x| \leq C_1 |x|^2, \text{ for some } C_1 > 0 \text{ for } |x| \leq 1/2,$$

$$(SM12.175) \quad |-\hat{s} \log(s/\hat{s}) + (s - \hat{s}) - \frac{s^2}{2\hat{s}}| \leq C_2 |s - \hat{s}|^3,$$

for some  $C_2 = C_2(s_*, \varepsilon) > 0$  and  $|s - \hat{s}| < \hat{s}/2, |\hat{s} - s_*| < \varepsilon$  for some fixed  $\varepsilon, s_* > 0$ .

Formulas (SM12.174), (SM12.175) describe the standard second order Taylor expansions of the logarithm in vicinity of  $x = 0$  and  $\hat{s} = s_*$ , respectively. Formula (SM12.173) can be proved via the following triangle-type inequality:

$$(SM12.176) \quad \begin{aligned} & |\Lambda_i - \widehat{\Lambda}_{sc,i}^t| \leq |\Lambda_i - \widetilde{\Lambda}_{b,app,i}^t| + |\widetilde{\Lambda}_{b,app,i}^t - \widetilde{\Lambda}_{b,i}^t| + |\widetilde{\Lambda}_{b,i}^t - \Lambda_i^*| + |\Lambda_i^* + \widehat{\Lambda}_{sc,i}^t|, \\ & \lambda \in C_{A,\delta}^t(\widetilde{\Lambda}_{b,app}^t). \end{aligned}$$

The first term in the right hand-side of (SM12.176) is of order  $o_{cp}(1)$  in view of the definition in (SM12.111) and the fact that  $\lambda \in C_{A,\delta}^t(\widetilde{\Lambda}_{b,app}^t)$  for some fixed  $\delta > 0$ . The last two terms are also  $o_{cp}(1)$  in view of Lemma SM12.5 and the fact that  $\widehat{\Lambda}_{sc,i}^t \rightarrow \Lambda_i^*$  a.s.  $Y^t, t \in [0, +\infty)$ . Finally, from (SM12.161) and again the fact that  $\widehat{\Lambda}_{sc,i}^t \rightarrow \Lambda_i^*$  a.s.  $Y^t, t \in (0, +\infty)$ , it follows that the second term in (SM12.176) is also of order  $o_{cp}(1)$ . This completes the proof of (SM12.173).

Using the restriction that  $\lambda \in C_{A,\delta}^t(\widetilde{\Lambda}_{b,app,t}^t)$  two first sums in (SM12.172) can be estimated as follows:

$$(SM12.177) \quad \begin{aligned} & \sum_{i \in I_1(\Lambda^*)} -tC_1 |\widetilde{\Lambda}_{b,i}^t - \widehat{\Lambda}_{sc,i}^t| \frac{|\Lambda_i - \widehat{\Lambda}_{sc,i}^t|^2}{|\widehat{\Lambda}_{sc,i}^t|^2} \geq \sum_{i \in I_1(\Lambda^*)} -tC_1 |\widetilde{\Lambda}_{b,i}^t - \widehat{\Lambda}_{sc,i}^t| \\ & \quad \times \left( \frac{2|\Lambda_i - \widetilde{\Lambda}_{b,app,i}^t|^2}{|\widehat{\Lambda}_{sc,i}^t|^2} + \frac{2|\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t|^2}{|\widehat{\Lambda}_{sc,i}^t|^2} \right) \\ & \geq \sum_{i \in I_1(\Lambda^*)} -tC_1 |\widetilde{\Lambda}_{b,i}^t - \widehat{\Lambda}_{sc,i}^t| \left( \frac{c\delta^2}{t|\widehat{\Lambda}_{sc,i}^t|^2} + \frac{2|\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t|^2}{|\widehat{\Lambda}_{sc,i}^t|^2} \right), \end{aligned}$$

where  $c$  depends only  $A$ . Using same argument for the second sum in (SM12.172) we obtain the following:

$$(SM12.178) \quad \begin{aligned} & \sum_{i \in I_1(\Lambda^*)} -tC_2 |\Lambda_i - \widehat{\Lambda}_{sc,i}^t|^3 \geq \sum_{i \in I_1(\Lambda^*)} -8tC_2 \left( |\Lambda_i - \widetilde{\Lambda}_{b,app,i}^t|^3 + |\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t|^3 \right) \\ & \geq \sum_{i \in I_1(\Lambda^*)} -8tC_2 \left( \frac{c\delta^3}{t^{3/2}} + |\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t|^3 \right), \end{aligned}$$

for  $\lambda \in C_{A,\delta}^t(\widetilde{\Lambda}_{b,app,t}^t)$ , where  $c$  depends only on  $A$ .

From (6.23), (SM12.117), and (SM12.122), the results of Lemmas SM12.12 and SM12.14 it follows that

$$(SM12.179) \quad t |\widetilde{\Lambda}_{b,i}^t - \widehat{\Lambda}_{sc,i}^t| \cdot |\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t|^2 = o_{cp}(1),$$

$$(SM12.180) \quad t |\widetilde{\Lambda}_{b,app,i}^t - \widehat{\Lambda}_{sc,i}^t|^3 = o_{cp}(1).$$

The above formulas imply that sums in (SM12.177), (SM12.178) are bounded from below and of order  $o_{cp}(1)$ .

The logarithmic term in (SM12.172) can be estimated as follows:

$$\begin{aligned}
\sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t\Lambda_i) &= \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t(\Lambda_i - \tilde{\Lambda}_{b,app,i}^t) + t\tilde{\Lambda}_{b,app,i}^t) \\
\text{(SM12.181)} \quad &\geq \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t|\Lambda_i - \tilde{\Lambda}_{b,app,i}^t| + t\tilde{\Lambda}_{b,app,i}^t) \\
&\geq \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(c\delta + t\tilde{\Lambda}_{b,app,i}^t), \lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t).
\end{aligned}$$

where  $c$  is some positive constant depending on  $A$ . Using (SM12.117), (SM12.122), and (SM12.124) from Lemma SM12.12 we obtain

$$\begin{aligned}
\text{(SM12.182)} \quad t\tilde{\Lambda}_{b,app,i}^t &= t(\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t) + t\hat{\Lambda}_{sc,i}^t \\
&= a_i^T \tilde{v}_{b,app}^t + t a_i^T \hat{\lambda}_{sc}^t \\
&= a_i^T (\tilde{v}_{b,app}^t - \tilde{v}^t) = o_{cp}(1), I_0(\Lambda^*)
\end{aligned}$$

Formulas (SM12.181), (SM12.182) imply that

$$\text{(SM12.183)} \quad \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t\Lambda_i) \geq \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(c\delta + o_{cp}(1)).$$

By choosing  $\delta$  smaller than some fixed constant (e.g.,  $\delta < c/2$ ) in (SM12.183) we find that the right hand-side in (SM12.183) becomes positive with conditional probability tending to one a.s.  $Y^t$ ,  $t \in [0, +\infty)$ . Therefore,

$$\text{(SM12.184)} \quad \sum_{i \in I_0(\Lambda^*)} -t\tilde{\Lambda}_{b,i}^t \log(t\Lambda_i) \geq o_{cp}(1), \lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t) \text{ for } \delta < c/2.$$

In addition, from the initial assumption in (6.23) it directly follows that

$$\text{(SM12.185)} \quad \sum_{i \in I_0(\Lambda^*)} t\hat{\Lambda}_{sc,i}^t \log(t\hat{\Lambda}_{sc,i}^t) - t\hat{\Lambda}_{sc,i}^t = o_{cp}(1).$$

Using (SM12.172), (SM12.177)–(SM12.180), (SM12.184), (SM12.185) we finally obtain:

$$\text{(SM12.186)} \quad \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - B^t(\lambda)] \geq o_{cp}(1) + \beta^t \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} (\varphi(\lambda) - \varphi(\hat{\lambda}_{sc}^t)).$$

Now, let us consider the third term in the left-hand side of (SM12.151). Using (SM12.118)–(SM12.120) we rewrite it as follows:

$$\begin{aligned}
\text{(SM12.187)} \quad B^t(\tilde{\lambda}_{b,app}^t) - B^t(\tilde{\lambda}_{app}^t) &= \tilde{B}^t(\tilde{\lambda}_{b,app}^t) - \tilde{B}^t(\tilde{\lambda}_{app}^t) \\
&\quad + \tilde{R}^t(\tilde{\lambda}_{b,app}^t) - \tilde{R}^t(\tilde{\lambda}_{app}^t).
\end{aligned}$$

From (SM12.108), the result of Lemma SM12.10, (SM12.118)–(SM12.120), (SM12.122), the result of Lemmas SM12.12 and SM12.14 and formula (SM12.187) it follows directly that

$$(SM12.188) \quad B^t(\tilde{\lambda}_{b,app}^t) - B^t(\tilde{\lambda}_{app}^t) = o_{cp}(1).$$

Now we estimate the last term in the right-hand side of (SM12.151). Using the same argument as in (SM12.172)–(SM12.185) one gets the following estimate:

$$(SM12.189) \quad \begin{aligned} B^t(\tilde{\lambda}_{app}^t) - A^t(\tilde{\lambda}_{app}^t) &\geq \sum_{i \in I_1(\Lambda^*)} -tC_1 |\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t| \frac{|\tilde{\Lambda}_{app,i}^t - \hat{\Lambda}_{sc,i}^t|^2}{|\hat{\Lambda}_{sc,i}^t|^2} \\ &+ \sum_{i \in I_1(\Lambda^*)} -tC_2 |\tilde{\Lambda}_{app,i}^t - \hat{\Lambda}_{sc,i}^t|^3 \\ &+ \sum_{i \in I_0(\Lambda^*)} t\tilde{\Lambda}_{b,i}^t \log(\tilde{\Lambda}_{app,i}^t) \\ &+ \sum_{i \in I_0(\Lambda^*)} -t\hat{\Lambda}_{sc,i}^t \log(t\hat{\Lambda}_{sc,i}^t) + t\hat{\Lambda}_{sc,i}^t \\ &- \beta^t(\varphi(\tilde{\lambda}_{app}^t) - \varphi(\hat{\lambda}_{sc}^t)). \end{aligned}$$

$$(SM12.190) \quad \begin{aligned} &\geq \sum_{i \in I_1(\Lambda^*)} -tC_1 |\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t| \frac{|\tilde{\Lambda}_{app,i}^t - \hat{\Lambda}_{sc,i}^t|^2}{|\hat{\Lambda}_{sc,i}^t|^2} \\ &+ \sum_{i \in I_1(\Lambda^*)} -tC_2 |\tilde{\Lambda}_{app,i}^t - \hat{\Lambda}_{sc,i}^t|^3 \\ &+ \sum_{i \in I_0(\Lambda^*)} t\tilde{\Lambda}_{b,i}^t \log(t\tilde{\Lambda}_{b,i}^t) \\ &+ \sum_{i \in I_0(\Lambda^*)} -t\hat{\Lambda}_{sc,i}^t \log(t\hat{\Lambda}_{sc,i}^t) + t\hat{\Lambda}_{sc,i}^t \\ &- \beta^t(\varphi(\tilde{\lambda}_{app}^t) - \varphi(\hat{\lambda}_{sc,i}^t)), \end{aligned}$$

where constants  $C_1, C_2$  depend only on  $A$ . To pass from (SM12.189) to (SM12.190) we have used the monotonicity of the logarithm (i.e.,  $\log(x+y) \geq \log(x)$ , for any  $y > 0$ ). The above estimate holds with conditional probability tending to one a.s.  $Y^t$ ,  $t \in [0, +\infty)$ .

From formulas (SM12.107), (SM12.109), (SM12.117), and (SM12.122), the results of Lemmas SM12.12 and SM12.14 it follows that

$$(SM12.191) \quad t |\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t| \cdot |\tilde{\Lambda}_{app,i}^t - \hat{\Lambda}_{sc,i}^t|^2 = o_{cp}(1),$$

$$(SM12.192) \quad t |\tilde{\Lambda}_{app,i}^t - \hat{\Lambda}_{sc,i}^t|^3 = o_{cp}(1).$$

In addition, using (SM12.147) in the proof of Lemma SM12.10 we find that

$$(SM12.193) \quad \sum_{i \in I_0(\Lambda^*)} t\tilde{\Lambda}_{b,i}^t \log(t\tilde{\Lambda}_{b,i}^t) = o_{cp}(1).$$

Putting together (SM12.190)–(SM12.193) and using again (SM12.185) we obtain

$$(SM12.194) \quad B^t(\tilde{\lambda}_{app}^t) - A^t(\tilde{\lambda}_{app}^t) \geq o_{cp}(1) - \beta^t(\varphi(\tilde{\lambda}_{app}^t) - \varphi(\tilde{\lambda}_{sc}^t)).$$

Formulas (SM12.151), (SM12.186), (SM12.188), and (SM12.194) imply that

$$(SM12.195) \quad \begin{aligned} \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [A^t(\lambda) - A^t(\tilde{\lambda}_{app}^t)] &= \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [B^t(\lambda) - B^t(\tilde{\lambda}_{b,app}^t)] \\ &+ \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} \beta^t(\varphi(\lambda) - \varphi(\tilde{\lambda}_{app}^t)) \\ &+ o_{cp}(1). \end{aligned}$$

**Lemma SM12.16.** *Let  $\beta^t, \varphi(\cdot)$  satisfy the assumptions of Theorem 6.10 and  $\tilde{\lambda}_{b,app}^t, \tilde{\lambda}_{app}^t$  be defined in (SM12.106), (SM12.107), respectively. Then,*

$$(SM12.196) \quad \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} \beta^t(\varphi(\lambda) - \varphi(\tilde{\lambda}_{app}^t)) = o_{cp}(1), \text{ a.s. } Y^t, t \in (0, +\infty).$$

Formula (SM12.152) directly follows from (SM12.195) and the result of Lemma SM12.16.

Lemma is proved. ■

### SM12.13. Proof of Lemma SM12.12.

*Proof.* To prove the claim we use essentially the same convexity argument as before, for example in Lemma SM12.11.

Let  $\delta > 0$  and

$$(SM12.197) \quad \tilde{\lambda}^t = \hat{\lambda}_{sc}^t + \frac{\tilde{u}^t}{\sqrt{t}} + \frac{\tilde{v}^t}{t} + \tilde{w}^t, \tilde{\lambda}^t \geq 0,$$

$$(SM12.198) \quad \lambda(u, v, w) = \hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + \frac{v}{t} + w, (u, v, w) \in \mathcal{U} \times \mathcal{V} \times \mathcal{W}, \lambda(u, v, w) \geq 0.$$

Recall that

$$(SM12.199) \quad \|u - \tilde{u}^t\|_2 + \|v - \tilde{v}^t\|_1 = \delta \text{ for } \lambda(u, v, w) \in C_{A,\delta}^t(\tilde{\lambda}^t).$$

where  $C_{A,\delta}^t(\cdot)$  is defined in (SM12.111).

Next we show that

$$(SM12.200) \quad P\left(\inf_{(u,v,w):\lambda(u,v,w) \in C_{A,\delta}^t(\tilde{\lambda}^t)} [B^t(u, v) - B^t(\tilde{u}^t, \tilde{v}^t)] > 0 \mid Y^t, t\right) \rightarrow 1 \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in [0, +\infty)$$

which together with the fact that  $\delta$  can be arbitrarily small and convexity of  $B^t(u, v)$ , implies the claim of the lemma. Using formulas (SM12.118)–(SM12.120) we obtain

$$(SM12.201) \quad \begin{aligned} B^t(u, v) - B^t(\tilde{u}^t, \tilde{v}^t) &= [\tilde{B}^t(u, v) - \tilde{B}^t(\tilde{u}^t, \tilde{v}^t)] \\ &+ [\tilde{R}^t(u, v) - \tilde{R}^t(\tilde{u}^t, \tilde{v}^t)], \\ &(u, v) \text{ s.t. } \exists w \in \mathcal{W}, \lambda(u, v, w) \in C_{A,\delta}^t(\tilde{\lambda}^t). \end{aligned}$$

From the facts that  $\tilde{\Lambda}_{b,i}^t \xrightarrow{c.p.} \Lambda_i^*$  (by Lemma SM12.5),  $\hat{\Lambda}_{sc,i}^t \xrightarrow{a.s.} \Lambda_i^*$  for  $i \in \{1, \dots, d\}$  (see (SM2.2), (6.22), and (6.23) in section SM2), the conditional tightness of  $\tilde{u}^t$  (Lemma SM12.14) and formulas (6.23), (SM12.122), (SM12.199) it follows that

$$(SM12.202) \quad \sup_{(u,v,w): \lambda(u,v,w) \in C_{A,\delta}^t(\tilde{\lambda}^t)} |\tilde{R}^t(u, v) - \tilde{R}^t(\tilde{u}^t, \tilde{v}^t)| = o_{cp}(1),$$

where  $\tilde{R}^t(\cdot)$  is defined in (SM12.120).

Formulas (SM12.201), (SM12.202) imply that

$$(SM12.203) \quad \inf_{(u,v,w): \lambda(u,v,w) \in C_{A,\delta}^t(\tilde{\lambda}^t)} [B^t(u, v) - B^t(\tilde{u}^t, \tilde{v}^t)] \geq \inf_{(u,v,w): \lambda(u,v,w) \in C_{A,\delta}^t(\tilde{\lambda}^t)} [\tilde{B}^t(u, v) - \tilde{B}^t(\tilde{u}^t, \tilde{v}^t)] + o_{cp}(1).$$

Since the positivity constraints in (SM12.125) include restrictions on  $u \in \mathcal{U}$  and also depend on  $w \in \mathcal{W}$ , for simplicity, we include  $w$  in the minimization problem as an independent variable

$$(SM12.204) \quad (\tilde{u}^t, \tilde{w}^t) = \arg \min_{\substack{(u,w): (1-\Pi_{\mathcal{V}})\tilde{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + w \geq 0 \\ u \in \mathcal{U}, w \in \mathcal{W}}} \sum_{i \in I_1(\Lambda^*)} -\sqrt{t}(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{a_i^T u}{\hat{\Lambda}_{sc,i}^t} + \frac{(a_i^T u)^2}{2\hat{\Lambda}_{sc,i}^t}.$$

Note that minimizer  $\tilde{u}^t$  in (SM12.204) coincides with the original solution from (SM12.125). The problem in (SM12.204) is convex and the strong duality is satisfied (e.g., by Slater's condition). From the Karush-Kuhn-Tucker necessary optimality conditions (see e.g., [SM3], Section 3.3) for the optimization problem in (SM12.204) and the strong duality it follows that

$$(SM12.205) \quad \exists \tilde{\mu}^t \succeq 0, \tilde{\mu}^t \in \mathcal{W}^\perp,$$

$$(SM12.206) \quad \sum_{i \in I_1(\Lambda^*)} -\sqrt{t}(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{\Pi_{\mathcal{U}} a_i}{\hat{\Lambda}_{sc,i}^t} + \frac{\Pi_{\mathcal{U}} a_i a_i^T \tilde{u}^t}{\hat{\Lambda}_{sc,i}^t} = \frac{\tilde{\mu}_{\mathcal{U}}^t}{\sqrt{t}}, \tilde{\mu}_{\mathcal{U}}^t = \Pi_{\mathcal{U}} \tilde{\mu}^t,$$

$$(SM12.207) \quad \tilde{\mu}_j^t \left( [(I - \Pi_{\mathcal{V}})\tilde{\lambda}_{sc}^t]_j + \frac{\tilde{u}_j^t}{\sqrt{t}} + \tilde{w}_j^t \right) = 0, j \in \{1, \dots, p\},$$

where  $(\tilde{u}^t, \tilde{w}^t)$  are defined in (SM12.204). Strong duality implies, in particular, that  $\tilde{\mu}^t$  is a solution for the dual problem and  $\tilde{\mu}^t \in \mathcal{W}^\perp$  (dual functional equals  $-\infty$  for  $\tilde{\mu}^t \notin \mathcal{W}^\perp$ ). Note also that the optimized functional in (SM12.204) is strongly convex in  $u$ , so  $\tilde{u}^t$  is always unique, whereas at least one  $\tilde{w}^t$  always exists, however, may not be unique. The latter fact does not pose any problem since the target functional is flat for  $w \in \mathcal{W}$ , so if not said otherwise, we choose any solution  $\tilde{w}^t$  in (SM12.204) so that positivity constraints are satisfied.

From (SM12.118), (SM12.205)-(SM12.207) it follows that

$$\begin{aligned}
\tilde{B}^t(u, v) - \tilde{B}^t(\tilde{u}^t, \tilde{v}^t) &= \sum_{i \in I_1(\Lambda^*)} -\sqrt{t} \frac{\tilde{\Lambda}_{b,i} - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i^T (u - \tilde{u}^t) + \frac{1}{2} \frac{(a_i^T u)^2 - (a_i^T \tilde{u}^t)^2}{\hat{\Lambda}_{sc,i}^t} \\
&+ \sum_{i \in I_0(\Lambda^*)} a_i^T (v - \tilde{v}^t) \\
&= \sum_{i \in I_1(\Lambda^*)} -\sqrt{t} \frac{\tilde{\Lambda}_{b,i} - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i^T (u - \tilde{u}^t) + \frac{1}{2} \frac{(a_i^T (u - \tilde{u}^t))^2}{\hat{\Lambda}_{sc,i}^t} \\
&+ \frac{(\tilde{u}^t)^T a_i a_i^T (u - \tilde{u}^t)}{\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_0(\Lambda^*)} a_i^T (v - \tilde{v}^t) \\
\text{(SM12.208)} \quad &= \left\langle \frac{\tilde{\mu}_{\mathcal{U}}^t}{\sqrt{t}}, u - \tilde{u}^t \right\rangle + \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \frac{|a_i^T (u - \tilde{u}^t)|^2}{\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_0(\Lambda^*)} a_i^T (v - \tilde{v}^t).
\end{aligned}$$

Note that

$$\text{(SM12.209)} \quad \left\langle \frac{\tilde{\mu}_{\mathcal{U}}^t}{\sqrt{t}}, u - \tilde{u}^t \right\rangle \geq 0,$$

$$\text{(SM12.210)} \quad v - \tilde{v}^t \succeq 0,$$

$$\text{for } (u, v) \in \mathcal{U} \times \mathcal{V} \text{ s.t. } \lambda(u, v, w) = \hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + \frac{v}{t} + w \succeq 0 \text{ for some } w \in \mathcal{W}.$$

Indeed, in view of (SM12.205), (SM12.207) the left hand-side in (SM12.209) can be rewritten as follows:

$$\begin{aligned}
\left\langle \frac{\tilde{\mu}_{\mathcal{U}}^t}{\sqrt{t}}, u - \tilde{u}^t \right\rangle &= \left\langle \tilde{\mu}_{\mathcal{U}}^t, \frac{u}{\sqrt{t}} - \frac{\tilde{u}^t}{\sqrt{t}} \right\rangle \\
\text{(SM12.211)} \quad &= \langle (I - \Pi_{\mathcal{V}}) \tilde{\mu}^t, (I - \Pi_{\mathcal{V}}) \hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} \rangle \\
&= \langle (I - \Pi_{\mathcal{V}}) \tilde{\mu}^t, \hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + \frac{v}{t} + w \rangle \\
&= \langle (I - \Pi_{\mathcal{V}}) \tilde{\mu}^t, \lambda(u, v, w) \rangle
\end{aligned}$$

Note also that from (SM12.205) and the definition of  $\mathcal{V}$  in (6.16) it follows that

$$\text{(SM12.212)} \quad \mu_{\mathcal{U}}^t = (I - \Pi_{\mathcal{V}}) \mu^t \succeq 0.$$

Formula (SM12.209) follows directly from (SM12.211) and (SM12.212), and the fact that  $\lambda(u, v, w) \succeq 0$ .

In turn, formula (SM12.210) follows from (SM12.122).

Formulas (SM12.199), (SM12.208)–(SM12.210) and the fact that  $\hat{\Lambda}_{sc,i} \rightarrow \Lambda_i^*$  for  $i \in \{1, \dots, d\}$  a.s.  $Y^t$ ,  $t \in [0, +\infty)$  (as a strongly consistent estimator), imply that with conditional probability tending to one a.s.  $Y^t$ ,  $t \in (0, +\infty)$  the following estimate holds:

$$\text{(SM12.213)} \quad \inf_{(u,v,w): \lambda(u,v,w) \in C_{A,\delta}^t(\lambda^t)} [\tilde{B}^t(u, v) - \tilde{B}^t(\tilde{u}^t, \tilde{v}^t)] \geq c\delta^2,$$



where  $c$  is some fixed positive constant depending only on  $\Lambda^*$  and  $A$ .

Formula (SM12.200) follows directly from (SM12.203), (SM12.213).

Lemma is proved. ■

**SM12.14. Proof of Lemma SM12.13.** Let  $\xi \in \mathbb{R}^{\#I_1(\Lambda^*)}$  be a parameter and consider  $\tilde{u}^t(\xi)$  defined in (SM12.127).

Since the positivity constraints in (SM12.127) include restrictions on  $u \in \mathcal{U}$  and  $w \in \mathcal{W}$ , for simplicity, we include  $w$  in the minimization problem as an independent variable

$$(SM12.214) \quad (\tilde{u}^t, \tilde{w}^t) = \underset{\substack{(u,w): (1-\Pi_{\mathcal{V}})\hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + w \succeq 0 \\ u \in \mathcal{U}, w \in \mathcal{W}}}{\arg \min} \quad -\xi^T C^t u + \frac{1}{2} u^T F^t u,$$

where

$$(SM12.215) \quad \begin{aligned} C^t &= (\hat{D}_{I_1(\Lambda^*)}^t)^{-1/2} A_{I_1(\Lambda^*)}, \quad F^t = \hat{F}_{I_1(\Lambda^*)}^t, \\ \hat{D}_{I_1(\Lambda^*)}^t, \hat{F}_{I_1(\Lambda^*)}^t &\text{ are defined in (SM12.128), (SM12.129).} \end{aligned}$$

The Lagrangian function for the primal problem in (SM12.214) is defined by the formula:

$$(SM12.216) \quad \mathcal{L}^t(u, w; \mu) = -\xi^T C^t u + \frac{1}{2} u^T F^t u - \mu^T \left( (1 - \Pi_{\mathcal{V}}) \hat{\lambda}_{sc}^t + \frac{u}{\sqrt{t}} + w \right),$$

$$(SM12.217) \quad u \in \mathcal{U}, w \in \mathcal{W}, \mu \succeq 0.$$

The dual function for  $G^t(\mu)$  and solution  $\mu^t$  for the dual problem are defined by the formulas:

$$(SM12.218) \quad G^t(\mu) = \inf_{u \in \mathcal{U}, w \in \mathcal{W}} \mathcal{L}^t(u, w; \mu), \quad \mu^t = \arg \max_{\mu \succeq 0} G^t(\mu).$$

From the Karush-Kuhn-Tucker necessary optimality conditions, the fact that the primal problem is strongly convex in  $u \in \mathcal{U}$  and the strong duality it follows that

$$(SM12.219) \quad \exists (u^t, w^t) \in \mathcal{U} \times \mathcal{W}, \mu^t \succeq 0, \mu^t \in \mathcal{W}^\perp \text{ s.t.}$$

$$(SM12.220) \quad (u^t, w^t) \text{ is a solution for the primal problem in (SM12.214),}$$

$$(SM12.221) \quad \mu^t = \mu^t(\xi) \text{ is a solution for the dual problem in (SM12.218),}$$

$$(SM12.222) \quad \nabla_{u,w} \mathcal{L}^t(u^t, w^t; \mu^t) = 0,$$

$$(SM12.223) \quad \left( (1 - \Pi_{\mathcal{V}}) \hat{\lambda}_{sc,j}^t + \frac{u_j^t}{\sqrt{t}} + w_j^t \right) \mu_j^t = 0, \quad j \in \{1, \dots, p\}.$$

Using (SM12.216), (SM12.222) we obtain the following:

$$(SM12.224) \quad -\Pi_{\mathcal{U}}(C^t)^T \xi + (\Pi_{\mathcal{U}} F^t \Pi_{\mathcal{U}}) u^t - \frac{\Pi_{\mathcal{U}} \mu^t(\xi)}{\sqrt{t}} = 0,$$

$$(SM12.225) \quad \Pi_{\mathcal{W}} \mu^t = 0,$$

where  $\Pi_{\mathcal{U}}, \Pi_{\mathcal{W}}$  are defined in (6.19). In what follows we use the following notations

$$(SM12.226) \quad C_{\mathcal{U}}^t = C^t \Pi_{\mathcal{U}}, \quad F_{\mathcal{U}}^t = (\Pi_{\mathcal{U}} F^t \Pi_{\mathcal{U}}), \quad \mu_{\mathcal{U}}^t = \Pi_{\mathcal{U}} \mu^t.$$

Strong consistency of  $\widehat{\lambda}_{sc}^t$  on  $\mathcal{U} \oplus \mathcal{V}$  and the Continuous Mapping Theorem imply that

$$(SM12.227) \quad C_{\mathcal{U}}^t \rightarrow C_{\mathcal{U}}^*, F_{\mathcal{U}}^t \rightarrow F_{\mathcal{U}}^* \text{ when } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty),$$

where

$$(SM12.228) \quad C_{\mathcal{U}}^* = \Pi_{\mathcal{U}} C^*, F_{\mathcal{U}}^* = \Pi_{\mathcal{U}} F^* \Pi_{\mathcal{U}},$$

$$(SM12.229) \quad C^* = (D_{I_1(\Lambda^*)}^*)^{-1/2} A_{I_1(\Lambda^*)}, D_{I_1(\Lambda^*)}^* = \text{diag}(\dots, \Lambda_i^*, \dots), i \in I_1(\Lambda^*),$$

$$(SM12.230) \quad F^* = \sum_{i \in I_1(\Lambda^*)} \frac{a_i a_i^T}{\Lambda_i^*} = A_{I_1(\Lambda^*)}^T (D_{I_1(\Lambda^*)}^*)^{-1} A_{I_1(\Lambda^*)}.$$

Using the notations from (SM12.226) formula (SM12.224) can be rewritten as follows:

$$(SM12.231) \quad u^t(\xi) = (F_{\mathcal{U}}^t)^{-1} (C_{\mathcal{U}}^t)^T \xi + (F_{\mathcal{U}}^t)^{-1} \frac{\mu_{\mathcal{U}}^t(\xi)}{\sqrt{t}}.$$

Note that  $F_{\mathcal{U}}^t$  is continuously invertible on  $\mathcal{U}$ , therefore  $(F_{\mathcal{U}}^t)^{-1}$  is well-defined. Moreover,  $(F_{\mathcal{U}}^t)^{-1} \rightarrow (F_{\mathcal{U}}^*)^{-1}$  for  $t \rightarrow +\infty$  a.s.  $Y^t, t \in [0, +\infty)$ . Next, we show that the following estimate always holds:

$$(SM12.232) \quad \left| \frac{\mu_{\mathcal{U}}^t(\xi)}{\sqrt{t}} \right| \leq 2 \max_{\sigma \in \sigma_{\mathcal{U}}(F_{\mathcal{U}}^t)} \sigma^{-1/2} \|(F_{\mathcal{U}}^t)^{-1/2}\| \|(C_{\mathcal{U}}^t)^T \xi\|,$$

where  $\sigma_{\mathcal{U}}(F_{\mathcal{U}}^t)$  denotes the spectrum of  $F_{\mathcal{U}}^t$  on  $\mathcal{U}$  (which in view of (SM12.227), (SM12.230) contains only non-zero positive elements starting from some  $t \geq t_0$ ).

We begin with characterization of mapping  $\mu_{\mathcal{U}}^t(\xi)$  via the dual problem in (SM12.218).

First, from (SM12.216) and (SM12.218) it follows that

$$(SM12.233) \quad G^t(\mu) = -\infty \text{ if } \mu \notin \mathcal{W}^{\perp}.$$

That is for  $\mu \notin \mathcal{W}^{\perp}$  the dual problem is unfeasible. In view of this and the strong duality, formulas in (SM12.218) can be rewritten as follows:

$$(SM12.234) \quad G^t(\mu) = \inf_{u \in \mathcal{U}} \mathcal{L}^t(u, 0; \mu), \mu \succeq 0, \mu \in \mathcal{W}^{\perp},$$

$$(SM12.235) \quad \mu^t = \arg \max_{\mu \succeq 0, \mu \in \mathcal{W}^{\perp}} G^t(\mu).$$

Using (SM12.216), (SM12.226), the first order optimality condition in (SM12.234) has the following form:

$$(SM12.236) \quad \begin{aligned} u_{min}^t(\mu) &= (F_{\mathcal{U}}^t)^{-1} (C_{\mathcal{U}}^t)^T \xi + (F_{\mathcal{U}}^t)^{-1} \frac{\mu_{\mathcal{U}}}{\sqrt{t}}, \\ \mu_{\mathcal{U}} &= \Pi_{\mathcal{U}} \mu, \mu \succeq 0, \mu \in \mathcal{W}^{\perp}. \end{aligned}$$

From (SM12.216), (SM12.218), (SM12.234), and (SM12.236) it follows that

$$\begin{aligned}
 (SM12.237) \quad G^t(\mu) &= \mathcal{L}^t(u_{min}^t(\mu), 0; \mu) = -\xi^T C_{\mathcal{U}}^t u_{min}^t(\mu) + \frac{1}{2} [u_{min}^t(\mu)]^T F_{\mathcal{U}}^t u_{min}^t(\mu) \\
 &\quad - \mu^T \left( (1 - \Pi_{\mathcal{V}}) \widehat{\lambda}_{sc}^t + \frac{u_{min}^t(\mu)}{\sqrt{t}} \right), \\
 &\quad \mu_{\mathcal{U}} = \Pi_{\mathcal{U}} \mu, \mu \succeq 0, \mu \in \mathcal{W}^{\perp}.
 \end{aligned}$$

Formulas (SM12.236), (SM12.237) imply that

$$\begin{aligned}
 (SM12.238) \quad G^t(\mu) &= -\frac{1}{2} \frac{\mu_{\mathcal{U}}^T}{\sqrt{t}} (F_{\mathcal{U}}^t)^{-1} \frac{\mu_{\mathcal{U}}}{\sqrt{t}} - \xi^T C_{\mathcal{U}}^t (F_{\mathcal{U}}^t)^{-1} \frac{\mu_{\mathcal{U}}}{\sqrt{t}} - \mu^T (I - \Pi_{\mathcal{V}}) \widehat{\lambda}_{sc}^t, \\
 &\quad \mu \succeq 0, \mu \in \mathcal{W}^{\perp}.
 \end{aligned}$$

From the facts that  $\mu \in \mathcal{W}^{\perp}$ ,  $\mu \succeq 0$  and the definition of  $\mathcal{V}$  in (6.16) it follows that

$$(SM12.239) \quad \mu_{\mathcal{U}} = (I - \Pi_{\mathcal{V}}) \mu = \begin{cases} \mu_j, & \text{if } \sum_{i \in I_0(\Lambda^*)} a_{ij} = 0, \\ 0, & \text{otherwise,} \end{cases} \Rightarrow \mu_{\mathcal{U}} = (I - \Pi_{\mathcal{V}}) \mu \succeq 0.$$

From (SM12.239) and the fact that  $\widehat{\lambda}_{sc}^t \succeq 0$  it follows that

$$(SM12.240) \quad \mu^T (I - \Pi_{\mathcal{V}}) \widehat{\lambda}_{sc}^t = [(I - \Pi_{\mathcal{V}}) \mu]^T \widehat{\lambda}_{sc}^t = \mu_{\mathcal{U}}^T \widehat{\lambda}_{sc}^t \geq 0.$$

From (SM12.238) one can see that minimizer  $\mu^t$  in (SM12.235) may not be unique, however, its projection  $\mu_{\mathcal{U}}^t$  is unique since functional  $G^t(\mu)$  is strongly convex in  $\mu_{\mathcal{U}}$ . At the same time, from (SM12.231) it follows that only  $\mu_{\mathcal{U}}^t$  is essential for  $\widetilde{u}^t(\xi)$ . In view of (SM12.231), (SM12.238), the optimization problem in (SM12.235) can be rewritten as follows:

$$(SM12.241) \quad \frac{\mu_{\mathcal{U}}^t}{\sqrt{t}} = \widetilde{\mu}_{\mathcal{U}}^t = \arg \min_{\mu_{\mathcal{U}} \in \Pi_{\mathcal{U}}(\mathbb{R}_+^p \cap \mathcal{W}^{\perp})} \frac{1}{2} \|(F_{\mathcal{U}}^t)^{-1/2} \mu_{\mathcal{U}} + (F_{\mathcal{U}}^t)^{-1/2} (C_{\mathcal{U}}^t)^T \xi\|^2 + \sqrt{t} \mu_{\mathcal{U}}^T \widehat{\lambda}_{sc}^t.$$

From (SM12.241) and the fact that  $0 \in \Pi_{\mathcal{U}}(\mathbb{R}_+^p \cap \mathcal{W}^{\perp})$  it follows that

$$(SM12.242) \quad \frac{1}{2} \|(F_{\mathcal{U}}^t)^{-1/2} \widetilde{\mu}_{\mathcal{U}}^t + (F_{\mathcal{U}}^t)^{-1/2} (C_{\mathcal{U}}^t)^T \xi\|^2 + \sqrt{t} \mu_{\mathcal{U}}^t \widehat{\lambda}_{sc}^t \leq \|(F_{\mathcal{U}}^t)^{-1/2} (C_{\mathcal{U}}^t)^T \xi\|^2,$$

where  $\widetilde{\mu}_{\mathcal{U}}^t$  is the solution in (SM12.241). Formulas (SM12.240), (SM12.242) imply that

$$(SM12.243) \quad |(F_{\mathcal{U}}^t)^{-1/2} \widetilde{\mu}_{\mathcal{U}}^t + (F_{\mathcal{U}}^t)^{-1/2} (C_{\mathcal{U}}^t)^T \xi| \leq \|(F_{\mathcal{U}}^t)^{-1/2} (C_{\mathcal{U}}^t)^T \xi\|.$$

which together with inequality  $|a + b| \geq |a| - |b|$  imply the following estimate

$$(SM12.244) \quad \|(F_{\mathcal{U}}^t)^{-1/2} \widetilde{\mu}_{\mathcal{U}}^t\| \leq 2 \|(F_{\mathcal{U}}^t)^{-1/2} (C_{\mathcal{U}}^t)^T \xi\|.$$

From (6.17), (SM12.226), (SM12.227), and (SM12.230) it follows that  $F_{\mathcal{U}}^t$  is of full rank on  $\mathcal{U}$  (starting from some  $t \geq t_0$  a.s.  $Y^t$ ,  $t \in [0, +\infty)$ ), therefore, for large  $t$  matrix  $(F_{\mathcal{U}}^t)^{-1/2}$  is positive definite, injective on  $\mathcal{U}$  and, hence,  $|(F_{\mathcal{U}}^t)^{-1/2} \widetilde{\mu}_{\mathcal{U}}^t| \geq \min_{\sigma \in \sigma_{\mathcal{U}}(F_{\mathcal{U}}^t)} \sigma^{1/2} \|\widetilde{\mu}_{\mathcal{U}}^t\|$ , where  $\sigma_{\mathcal{U}}(\cdot)$  denotes the spectrum of an operator acting on  $\mathcal{U}$ .

The above argument with formula (SM12.244) directly imply (SM12.232).

Formulas (SM12.130)–(SM12.135) follow from (SM12.226)–(SM12.230), (SM12.231) and (SM12.232).

Lemma is proved.

**SM12.15. Proof of Lemma SM12.14.**

*Proof.* In view of Algorithm 4 intensities  $\tilde{\Lambda}_{b,i}^t$  can be represented as follows:

$$(SM12.245) \quad \tilde{\Lambda}_{b,i}^t = \frac{1}{\theta^t + t} \sum_{k=1}^{Y_i^t} w_{ik} + \tilde{r}_{b,\mathcal{M},i}^t, \quad i \in I_1(\Lambda^*),$$

$$(SM12.246) \quad \{w_{ik}\}_{k=1, i=1}^{\infty, d} \text{ are mutually independent, } w_{ik} \sim \Gamma(1, 1),$$

where

$$(SM12.247) \quad \begin{aligned} \tilde{r}_{b,\mathcal{M},i}^t &| \tilde{\Lambda}_{\mathcal{M},i}^t, Y^t, t \sim \Gamma(\theta^t \Lambda_{\mathcal{M},i}^t, (\theta^t + t)^{-1}), \\ \tilde{\Lambda}_{\mathcal{M},i}^t &\text{ are sampled in Algorithm 3.} \end{aligned}$$

In particular,

$$(SM12.248) \quad \sqrt{t} r_{b,\mathcal{M},i}^t = o_{cp}(1).$$

Indeed, from (SM12.72), (SM12.247) and the Markov inequality it holds that

$$(SM12.249) \quad \begin{aligned} P(\sqrt{t} r_{b,\mathcal{M},i}^t > \delta \mid Y^t, t) &\leq \frac{\sqrt{t} \theta^t}{\delta(\theta^t + t)} E[\tilde{\Lambda}_{\mathcal{M},i}^t \mid Y^t, t] \\ &\leq \frac{\sqrt{t} \theta^t}{\delta(\theta^t + t)} \sum_{i \in I_1(\Lambda^*)} \frac{Y_i^t}{t} \rightarrow 0 \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty), \end{aligned}$$

where  $\delta$  is arbitrary positive value.

Using the Central Limit Theorem for sums of  $w_{ik}$  in (SM12.245), (SM12.246) and the Strong Law of Large Numbers for  $Y^t$  (see Theorem SM2.1, formula (SM2.2)) and the fact that  $\theta^t = o(\sqrt{t})$ , we obtain:

$$(SM12.250) \quad \frac{\sqrt{t}}{(\theta^t + t) \sqrt{Y_i^t/t}} \sum_{k=1}^{Y_i^t} (w_{ik} - 1) \xrightarrow{c.d.} \mathcal{N}(0, 1).$$

Due to mutual independence between  $w_{ik}$ , the above convergence holds for all components  $i \in I_1(\Lambda^*)$ , hence, as for the vector in  $\mathbb{R}^{\#I_1(\Lambda^*)}$ .

Using formula (SM12.136) we obtain:

$$(SM12.251) \quad \begin{aligned} A_{I_1(\Lambda^*)}^T (\hat{D}_{I_1(\Lambda^*)}^t)^{-1/2} \tilde{\xi}^t &= \sum_{i \in I_1(\Lambda^*)} \sqrt{t} \frac{\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i \\ &= \sum_{i \in I_1(\Lambda^*)} \sqrt{t} \frac{\tilde{\Lambda}_{b,i}^t - Y_i^t/t}{\hat{\Lambda}_{sc,i}^t} a_i + \sum_{i \in I_1(\Lambda^*)} \sqrt{t} \frac{Y_i^t/t - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i. \end{aligned}$$

The first sum is conditionally tight in view of the Prokhorov theorem on tightness of weakly convergence sequences and the result in (SM12.250). Due to (6.22) the second sum is simply bounded for large  $t$  for almost any trajectory  $Y^t, t \in (0, +\infty)$ . These arguments directly imply conditional tightness of  $A_{I_1(\Lambda^*)}^T (\hat{D}_{I_1(\Lambda^*)}^t)^{-1/2} \tilde{\xi}^t$  for almost any trajectory  $Y^t, t \in [0, +\infty)$ .

Lemma is proved. ■

**SM12.16. Proof of Lemma SM12.15.**

*Proof.* Since  $B^t(\lambda)$  is proportional to  $t$  in (SM12.105), it suffices to prove (SM12.153) for normalized process  $B^t(\lambda)/t$  which we denote here by  $G^t(\lambda)$ , that is

$$(SM12.252) \quad G^t(\lambda) = \sum_{i \in I_1(\Lambda^*)} -(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{\Lambda_i - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} + \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \frac{(\Lambda_i - \hat{\Lambda}_{sc,i}^t)^2}{\hat{\Lambda}_{sc,i}^t} \\ + \sum_{i \in I_0(\Lambda^*)} \Lambda_i, \quad \Lambda_i = a_i^T \lambda, \quad i \in \{1, \dots, d\}.$$

Note also that minimizers of  $B^t$  and of  $G^t$  coincide.

From the necessary Karush-Kuhn-Tucker optimality conditions in (SM12.106) (see e.g., [SM3], Section 3.3) it follows that

$$(SM12.253) \quad \exists \tilde{\lambda}_{b,app}^t, \tilde{\mu}_{b,app}^t \in \mathbb{R}_+^p \text{ such that} \\ - \sum_{i \in I_1(\Lambda^*)} \frac{\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i + \sum_{i \in I_1(\Lambda^*)} \frac{\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} a_i + \sum_{i \in I_0(\Lambda^*)} a_i - \tilde{\mu}_{b,app}^t = 0,$$

$$(SM12.254) \quad \tilde{\Lambda}_{b,app}^t = A \tilde{\lambda}_{b,app}^t, \\ \tilde{\mu}_{b,app,j}^t \tilde{\lambda}_{b,app,j}^t = 0 \text{ for all } j \in \{1, \dots, p\}.$$

Multiplying both sides of (SM12.253) on  $(\tilde{\lambda}_{b,app}^t - \hat{\lambda}_{sc}^t)$  and using formula (SM12.254) we obtain following formulas:

$$(SM12.255) \quad - \langle \tilde{\mu}_{b,app}^t, \hat{\lambda}_{sc}^t \rangle = - \sum_{i \in I_1(\Lambda^*)} \frac{(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t)(\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t)}{\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_1(\Lambda^*)} \frac{(\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t)^2}{\hat{\Lambda}_{sc,i}^t}$$

$$(SM12.256) \quad + \sum_{i \in I_0(\Lambda^*)} \tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t,$$

$$(SM12.257) \quad - \langle \tilde{\mu}_{b,app}^t, \hat{\lambda}_{sc}^t \rangle = \sum_{i \in I_1(\Lambda^*)} \tilde{\Lambda}_{b,i}^t - \tilde{\Lambda}_{b,app,i}^t - \sum_{i \in I_0(\Lambda^*)} \hat{\Lambda}_{sc,i}^t.$$

From formulas (SM12.252), (SM12.253), and (SM12.255) it follows that

$$(SM12.257) \quad G^t(\tilde{\lambda}_{b,app}^t) = - \langle \tilde{\mu}_{b,app}^t, \hat{\lambda}_{sc}^t \rangle - \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \frac{(\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t)^2}{\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_0(\Lambda^*)} \hat{\Lambda}_{sc,i}^t.$$

Using (SM12.252)–(SM12.257) we get the following identity:

(SM12.258)

$$\begin{aligned}
G^t(\lambda) - G^t(\tilde{\lambda}_{b,app}^t) &= \sum_{i \in I_1(\Lambda^*)} -(\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{\Lambda_i - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_0(\Lambda^*)} \Lambda_i - \hat{\Lambda}_{sc,i}^t \\
&+ \frac{1}{2} \sum_{i \in I_1(\Lambda^*)} \frac{(\Lambda_i - \hat{\Lambda}_{sc,i}^t)^2 + (\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t)^2}{\hat{\Lambda}_{sc,i}^t} + \langle \tilde{\mu}_{b,app}^t, \hat{\lambda}_{sc}^t \rangle \\
&= \sum_{i \in I_1(\Lambda^*)} \frac{(\Lambda_i - \tilde{\Lambda}_{b,app,i}^t)^2}{2\hat{\Lambda}_{sc,i}^t} + \sum_{i \in I_1(\Lambda^*)} \frac{(\tilde{\Lambda}_{b,app,i}^t - \hat{\Lambda}_{sc,i}^t)(\Lambda_i - \hat{\Lambda}_{sc,i}^t)}{\hat{\Lambda}_{sc,i}^t} \\
&+ \sum_{i \in I_0(\Lambda^*)} \Lambda_i - \hat{\Lambda}_{sc,i}^t + \sum_{i \in I_1(\Lambda^*)} \tilde{\Lambda}_{b,app,i}^t - \tilde{\Lambda}_{b,i}^t + \sum_{i \in I_0(\Lambda^*)} \hat{\Lambda}_{sc,i}^t \\
&- \sum_{i \in I_1(\Lambda^*)} (\tilde{\Lambda}_{b,i}^t - \hat{\Lambda}_{sc,i}^t) \frac{\Lambda_i - \hat{\Lambda}_{sc,i}^t}{\hat{\Lambda}_{sc,i}^t} \\
&= \sum_{i \in I_1(\Lambda^*)} \frac{(\Lambda_i - \tilde{\Lambda}_{b,app,i}^t)^2}{2\Lambda_i^*} + \sum_{i \in I_0(\Lambda^*)} \Lambda_i + \sum_{i \in I_1(\Lambda^*)} \frac{\tilde{\Lambda}_{b,app,i}^t - \tilde{\Lambda}_{b,i}^t}{\hat{\Lambda}_{sc,i}^t} \Lambda_i.
\end{aligned}$$

Formulas (SM12.153)–(SM12.155) follow from (SM12.252)–(SM12.254) and (SM12.258).

Lemma is proved. ■

### SM12.17. Proof of Lemma SM12.16.

*Proof.* Consider the following formula

$$(SM12.259) \quad \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [\varphi(\lambda) - \varphi(\tilde{\lambda}_{app}^t)] = \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} [\varphi(\lambda - \varphi(\tilde{\lambda}_{b,app}^t))] + [\varphi(\tilde{\lambda}_{b,app}^t) - \varphi(\tilde{\lambda}_{app}^t)].$$

Recall that  $\tilde{\lambda}_{b,app}^t$  may not be chosen uniquely since the functional  $B^t(\lambda)$  is strongly convex only in directions from  $\text{Span}\{a_i : i \in I_1(\Lambda^*)\}$  (see formula (SM12.105)) and it is flat in directions from  $\ker A$ . From the strong convexity of  $B^t(\lambda)$  on  $\text{Span}\{a_i : i \in I_1(\Lambda^*)\}$  and formulas (SM12.105), (SM12.106), (SM12.117) it follows that  $\tilde{u}_{b,app}^t = \sqrt{t}\Pi_{\mathcal{U}}(\tilde{\lambda}_{b,app}^t - \hat{\lambda}_{sc}^t)$  is unique. At the same time, from (6.23), (SM12.122) and the result of Lemma SM12.12 it follows that

$$(SM12.260) \quad \tilde{v}_{b,app}^t = t\Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \hat{\lambda}_{sc}^t) = o_{cp}(1),$$

where the above formula is understood as a uniform bound on the set of all possible minimizers  $\tilde{\lambda}_{b,app}^t$ . We may assume that for each  $t$  there is some unique  $\tilde{v}_{b,app}^t$ .

Then, to choose uniquely  $\tilde{\lambda}_{b,app}^t$  one has to fix its projection onto  $\mathcal{W}$  regarding the positivity constraints. Consider the following mapping

$$(SM12.261) \quad \begin{aligned} w(u, v) &= \arg \min_{\substack{w: \lambda_* + u + v + w \geq 0 \\ w \in \mathcal{W}}} \varphi(\lambda_* + u + v + w), \\ u &\in \mathcal{U}, v \in \mathcal{V} : (\lambda_* + u + v + \mathcal{W}) \cap \mathbb{R}_+^p \neq \emptyset, \end{aligned}$$

where  $\lambda_*$  is the true parameter. From the strict convexity of  $\varphi(\cdot)$  along  $\ker A$  (by the assumption in (2.12)), the definition of  $\mathcal{W}$  in (6.18) and the result of Lemma SM12.1 it follows that  $w(u, v)$  is one-to-one and continuous in  $(u, v)$  on its domain of definition.

Note that

$$(SM12.262) \quad w_* = w(0, 0) = w_{A, \lambda_*}(0, 0),$$

where  $w_{A, \lambda}(\cdot, \cdot)$  is defined in (SM12.261) ( $w_{A, \lambda_*}(0, 0)$  appears in Theorems 6.4 and 6.5). The property that  $w_* \in \mathcal{W}$  can be proved by the contradiction argument. Assume that  $w_* \in \ker A$  but  $w_* \notin \mathcal{W}$  and  $w_* \neq 0$ . Then, from the definition of  $\mathcal{V}, \mathcal{U}, \mathcal{W}$  it follows that

$$(SM12.263) \quad \exists i \in I_0(\Lambda^*), j \in \{1, \dots, p\} : a_{ij} > 0, w_{*j} > 0.$$

At the same time from the fact that  $w_* \in \ker A$  it follows that

$$(SM12.264) \quad 0 = \sum_{i \in I_0(\Lambda^*)} a_i^T w_* = \sum_{j=1}^p \left( \sum_{i \in I_0(\Lambda^*)} a_{ij} \right) w_{*j}$$

Formulas (SM12.263), (SM12.264) imply that

$$(SM12.265) \quad \exists i' \in I_0(\Lambda^*), j' \in \{1, \dots, p\} : a_{i'j'} > 0, w_{*j'} < 0.$$

At the same time, from the definition of  $I_0(\Lambda^*)$  in (2.2) it follows that  $\lambda_{*j'} = 0$  which together with the results from (SM12.265) contradicts the positivity constraint in (SM12.262). Thus,  $w_* \in \mathcal{W}$ .

Let

$$(SM12.266) \quad \tilde{w}_{b, app}^t = w \left( \Pi_{\mathcal{U}}(\hat{\lambda}_{sc}^t - \lambda_*) + \frac{\tilde{u}_{b, app}^t}{\sqrt{t}}, \Pi_{\mathcal{V}}(\hat{\lambda}_{sc}^t - \lambda_*) + \frac{\tilde{v}_{b, app}^t}{t} \right),$$

where  $\tilde{u}_{b, app}^t, \tilde{v}_{b, app}^t$  are defined in (SM12.106), (SM12.117),  $w$  is the mapping from (SM12.261). Recall that  $\tilde{\lambda}_{b, app}^t$  from (SM12.106) can be rewritten via the parametrization in (SM12.117) as follows

$$(SM12.267) \quad \tilde{\lambda}_{b, app}^t = \hat{\lambda}_{sc}^t + \frac{\tilde{u}_{b, app}^t}{\sqrt{t}} + \frac{\tilde{v}_{b, app}^t}{t} + \tilde{w}_{b, app}^t,$$

where  $\tilde{w}_{b, app}^t$  is chosen in (SM12.266). For  $\tilde{\lambda}_{b, app}^t$  from (SM12.267) it holds that

$$(SM12.268) \quad \tilde{\lambda}_{b, app}^t \xrightarrow{c.p.} \lambda_* + w_* \text{ for } t \rightarrow +\infty, \text{ a.s. } Y^t, t \in (0, +\infty),$$

where  $w_*$  is defined in (SM12.262).

Indeed, formula (SM12.268) follows from the fact that  $\Pi_{\mathcal{U} \oplus \mathcal{V}} \hat{\lambda}_{sc}^t \xrightarrow{c.p.} \Pi_{\mathcal{U} \oplus \mathcal{V}} \lambda_*$ , the fact that  $\tilde{u}_{b, app}^t / \sqrt{t} = o_{cp}(1)$ ,  $\tilde{v}_{b, app}^t / t = o_{cp}(1)$  (see formula (SM12.122) and results of Lemma SM12.14) and the continuity of mapping  $w$ .

From the local Lipschitz continuity of  $\varphi$  and (SM12.108), (SM12.109), (SM12.268) it follows that there exists some universal constant  $L > 0$  such that with conditional probability tending to one a.s.  $Y^t$ ,  $t \in (0, +\infty)$  it holds that:

$$(SM12.269) \quad \varphi(\tilde{\lambda}_{b,app}^t) - \varphi(\tilde{\lambda}_{app}^t) \leq L \|\tilde{\lambda}_{b,app}^t - \tilde{\lambda}_{app}^t\|.$$

In particular, from (SM12.108), (SM12.109), (SM12.269) it follows that

$$(SM12.270) \quad \beta^t(\varphi(\tilde{\lambda}_{b,app}^t) - \varphi(\tilde{\lambda}_{app}^t)) = o_{cp}(1).$$

It is left to show that the first term in (SM12.259) is also of order  $o_{cp}(1)$ . For this we use extensively the results from [SM33] on the Lipschitz-continuity of inf-projections.

The first term in (SM12.259) can be rewritten as taking the infimum two times:

$$(SM12.271) \quad \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} (\varphi(\lambda) - \varphi(\tilde{\lambda}_{b,app}^t)) = \inf_{\substack{(u,v) \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t) \\ (u,v) \in \mathcal{U} \times \mathcal{V}}} [\varphi_*(\Pi_{\mathcal{U}}(\tilde{\lambda}_{b,app}^t - \lambda_*) + \frac{u}{\sqrt{t}}, \Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \lambda_*) + \frac{v}{t}) \\ - \varphi_*(\Pi_{\mathcal{U}}(\tilde{\lambda}_{b,app}^t - \lambda_*), \Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \lambda_*))],$$

where

$$(SM12.272) \quad \varphi_*(u, v) = \inf_{\substack{w: \lambda_* + u + v + w \succeq 0, \\ w \in \mathcal{W}}} \varphi(\lambda_* + u + v + w), \\ u \in \mathcal{U}, v \in \mathcal{V} : (\lambda_* + u + v + \mathcal{W}) \cap \mathbb{R}_+^p \neq \emptyset.$$

The expression in the square brackets in (SM12.271) is essentially the variation of the inf-projection for  $\varphi_*(u, v)$  for parameter  $(u, v) \in \mathcal{U} \times \mathcal{V}$  in the vicinity of zero along  $\mathcal{U} \oplus \mathcal{V}$ . Indeed, this follows from the facts that  $\Pi_{\mathcal{U}}(\tilde{\lambda}_{b,app}^t - \lambda_*)$  and  $\Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \lambda_*)$  are both of order  $o_{cp}(1)$  and  $u/\sqrt{t}$ ,  $v/t$  are also  $o_{cp}(1)$  in view of the fact that  $(u, v) \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)$ .

Using Theorem 3.4 and examples in Section 4 (pp. 278-282) of [SM33] we find that  $\varphi_*(u, v)$  is locally Lipschitz continuous.

Indeed, consider the optimization problem in (SM12.272), where  $(u, v) \in \mathcal{U} \times \mathcal{V}$  is a parameter. Then, the problem can be rewritten as follows:

$$(SM12.273) \quad \inf_w \varphi_0((u, v); w), \quad \varphi_0 : (\mathcal{U} \times \mathcal{V}) \times \mathcal{W} \rightarrow \overline{\mathbb{R}},$$

$$(SM12.274) \quad \varphi_0((u, v); w) = \begin{cases} \varphi(\lambda_* + u + v + w), & \text{if } \lambda_* + u + v + w \succeq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

where  $\overline{\mathbb{R}}$  denotes the extended real line. From the fact that  $\varphi(\cdot)$  is locally Lipschitz continuous it is easy to see that  $\varphi_0$  is locally Lipschitz continuous on  $D = \{(u, v, w) \in \mathcal{U} \times \mathcal{V} \times \mathcal{W} : \lambda_* + u + v + w \succeq 0\}$ , where the latter is a polyhedral subset of  $\mathcal{U} \times \mathcal{V} \times \mathcal{W}$ .

Consider the feasibility mapping

$$(SM12.275) \quad S : \mathcal{U} \times \mathcal{V} \rightrightarrows \mathcal{W} \text{ with } S(u, v) = \{w \in \mathcal{W} : \lambda_* + u + v + w \succeq 0\},$$



where  $\rightrightarrows$  denotes the property to be a set-valued mapping. From (SM12.275) one can see that  $\text{gph } S = D$  (gph denotes the graph of a mapping). Therefore,  $\text{gph } S$  is polyhedral and, hence, the Proposition 4.1 from [SM33] applies to our case (see also Example 9.35 in [SM27]), so mapping  $S$  in (SM12.275) is Lipschitz continuous on  $\text{dom } S$  (as set-valued mapping). At the same time, the result of Lemma SM12.2 implies that feasibility mapping  $S$  is locally bounded which yields level boundedness in  $w$  locally uniformly in  $(u, v)$  of  $\varphi_0(\cdot, \cdot)$ . The above properties are exactly the same as in Section 4 of [SM33], so Theorem 3.4 therein applies to the case of  $\varphi_0$  from (SM12.273) and  $\varphi_*(u, v) = \inf_w \varphi((u, v); w)$  is locally Lipschitz continuous.

Hence, there exists a constant  $L > 0$  such that with conditional probability tending to one a.s.  $Y^t$ ,  $t \in (0, +\infty)$  the following holds

$$\begin{aligned} & \left| \varphi_*(\Pi_{\mathcal{U}}(\tilde{\lambda}_{b,app}^t - \lambda_*) + \frac{u}{\sqrt{t}}, \Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \lambda_*) + \frac{v}{t}) - \varphi_*(\Pi_{\mathcal{U}}(\tilde{\lambda}_{b,app}^t - \lambda_*), \Pi_{\mathcal{V}}(\tilde{\lambda}_{b,app}^t - \lambda_*)) \right| \\ \text{(SM12.276)} \quad & \leq L \left( \frac{\|u\|}{\sqrt{t}} + \frac{\|v\|}{t} \right) \leq L \left( \frac{\delta}{\sqrt{t}} + c \frac{\delta}{t} \right) \text{ for any } (u, v) \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t), \end{aligned}$$

where  $c$  is a positive constant depending only on dimension  $p$ . Using formulas (SM12.271), (SM12.276) and the assumption that  $\beta^t = o(\sqrt{t})$  we obtain

$$\text{(SM12.277)} \quad \beta^t \cdot \inf_{\lambda \in C_{A,\delta}^t(\tilde{\lambda}_{b,app}^t)} (\varphi(\lambda) - \varphi(\tilde{\lambda}_{b,app}^t)) = o_{cp}(1).$$

Formula (SM12.196) directly follows from (SM12.270) and (SM12.277).

Lemma is proved. ■

## REFERENCES

- [1] H. ATTOUCH AND R. J.-B. WETS, *Quantitative stability of variational systems. ii. a framework for nonlinear conditioning*, SIAM Journal on Optimization, 3 (1993), pp. 359–381, <https://doi.org/10.1137/0803016>.
- [2] R. G. AYKROYD AND P. J. GREEN, *Global and local priors, and the location of lesions using gamma-camera imagery*, Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences, 337 (1991), pp. 323–342.
- [3] D. P. BERTSEKAS, *Nonlinear programming*, Journal of Operational Research Society, 48 (1997), p. 334.
- [4] N. A. BOCHKINA AND P. J. GREEN, *The Bernstein–von Mises theorem and nonregular models*, The Annals of Statistics, 42 (2014), pp. 1850–1878.
- [5] D. J. DALEY AND D. VERE-JONES, *An introduction to the theory of point processes: volume II: general theory and structure*, Springer Science & Business Media, 2007.
- [6] A. DE PIERRO, *On the relation between the ISRA and the EM algorithm for positron emission tomography*, IEEE Transactions on Medical Imaging, 12 (1993), pp. 328–333.
- [7] H. ERDOGAN AND J. FESSLER, *Monotonic algorithms for transmission tomography*, IEEE Transactions on Medical Imaging, 18 (1999), pp. 801–814, <https://doi.org/10.1109/42.802758>.
- [8] J. FESSLER AND A. HERO, *Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms*, IEEE Transactions on Image Processing, 4 (1995), pp. 1417–1429.
- [9] M. FILIPOVIĆ, T. DAUTREMER, C. COMTAT, S. STUTE, AND E. BARAT, *Reconstruction, analysis and interpretation of posterior probability distributions of pet images, using the posterior bootstrap*, Physics in Medicine & Biology, 66 (2021), p. 125018.

- [10] E. FONG, S. LYDDON, AND C. HOLMES, *Scalable nonparametric sampling from multimodal posteriors with the posterior bootstrap*, in Proceedings of the 36th International Conference on Machine Learning, vol. 97, PMLR, 09–15 Jun 2019, pp. 1952–1962.
- [11] F. GONCHAROV, *Weighted Radon transforms and their applications*, PhD thesis, Université Paris Saclay (COmUE), 2019.
- [12] P. J. GREEN, *Bayesian reconstructions from emission tomography data using a modified em algorithm*, IEEE Trans. Med. Imag., 9 (1990), pp. 84–93.
- [13] A. GUT, *Probability: a graduate course*, New York, NY: Springer, 2013, <https://doi.org/10.1007/978-1-4614-4708-5>.
- [14] G. HAN, Z. LIANG, AND J. YOU, *A fast ray-tracing technique for tct and ect studies*, in 1999 IEEE Nuclear Science Symposium. Conference Record. 1999 Nuclear Science Symposium and Medical Imaging Conference (Cat. No.99CH37019), vol. 3, 1999, pp. 1515–1518 vol.3, <https://doi.org/10.1109/NSSMIC.1999.842846>.
- [15] J. HERMANS, A. DELAUNOY, F. ROZET, A. WEHENKEL, AND G. LOUPPE, *Averting a crisis in simulation-based inference*, arXiv preprint arXiv:2110.06581v2, (2021).
- [16] N. L. HJORT AND D. POLLARD, *Asymptotics for minimisers of convex processes*, arXiv preprint arXiv:1107.3806, (2011).
- [17] K. LANGE, D. R. HUNTER, AND I. YANG, *Optimization transfer using surrogate objective functions*, Journal of Computational and Graphical Statistics, 9 (2000), pp. 1–20, <https://doi.org/10.1080/10618600.2000.10474858>.
- [18] J. S. LIU, W. H. WONG, AND A. KONG, *Covariance structure of the gibbs sampler with applications to the comparisons of estimators and augmentation schemes*, Biometrika, 81 (1994), pp. 27–40.
- [19] A. Y. LO, *Bayesian nonparametric statistical inference for poisson point processes*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 59 (1982), pp. 55–66.
- [20] S. LYDDON, S. WALKER, AND C. HOLMES, *Nonparametric learning from bayesian models with randomized objective functions*, Advances in Neural Information Processing Systems, (2018).
- [21] F. NATTERER, *The mathematics of computerized tomography*, Society for Industrial and Applied Mathematics, 2001.
- [22] M. A. NEWTON AND A. E. RAFTERY, *Approximate bayesian inference with the weighted likelihood bootstrap*, Journal of the Royal Statistical Society: Series B, 56 (1994), pp. 3–26.
- [23] T. L. NG AND M. A. NEWTON, *Random weighting in LASSO regression*, Electronic Journal of Statistics, 16 (2022), pp. 3430 – 3481.
- [24] R. NOVIKOV, *5. non-abelian radon transform and its applications*, in The Radon Transform, De Gruyter, 2019, pp. 115–128.
- [25] E. POMPE, *Introducing prior information in weighted likelihood bootstrap with applications to model misspecification*, arXiv preprint arXiv:2103.14445, (2021).
- [26] E. T. QUINTO, *The invertibility of rotation invariant radon transforms*, Journal of Mathematical Analysis and Applications, 91 (1983), pp. 510–522.
- [27] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational analysis*, vol. 317, Springer Science & Business Media, 2009, <https://doi.org/10.1007/978-3-642-02431-3>.
- [28] R. SIDDON, *Fast calculation of the exact radiological path for a three-dimensional ct array.*, Medical physics, 12 2 (1985), pp. 252–5.
- [29] A. W. VAN DER VAART, *Asymptotic statistics*, vol. 3, Cambridge university press, 2000.
- [30] F. VASCONCELOS, B. HE, N. SINGH, AND Y. W. TEH, *Uncertainr: Uncertainty quantification of end-to-end implicit neural representations for computed tomography*, arXiv preprint arXiv:2202.10847v2, (2022).
- [31] D. W. WALKUP AND R. J.-B. WETS, *A lipschitzian characterization of convex polyhedra*, Proceedings of the American Mathematical Society, (1969), pp. 167–173.
- [32] G. WANG AND J. QI, *Edge-preserving pet image reconstruction using trust optimization transfer*, IEEE Transactions on Medical Imaging, 34 (2015), pp. 930–939, <https://doi.org/10.1109/TMI.2014.2371392>.
- [33] R. J.-B. WETS, *Lipschitz continuity of inf-projections*, Computational Optimization and Applications, 25 (2003), pp. 269–282, <https://doi.org/10.1023/A:1022925725741>.