

Supplemental Appendix to “A Variational View on Statistical Multiscale Estimation”

We start with the relation between different variational formulations, and then provide the proofs of the assertions in the article.

A. RELATIONS BETWEEN CONSTRAINED AND UNCONSTRAINED MINIMIZATION

There are at least two close connections between the constrained optimization in Equation 5 and the unconstrained optimization in Equation 4. The first one arises from a Lagrangian multiplier approach, whereas the second arises from Fenchel duality. In what follows we analyze both approaches.

A.1. Lagrangian Multiplier Approach

Let $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ and $G: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be convex proper functionals, and let $\mathbf{X} \in \mathbb{R}^{n \times p}$. It is convenient to allow a convex functional to attain the value ∞ . The set $\mathcal{D}(R) := \{\beta \in \mathbb{R}^p: R(\beta) < \infty\}$ where the functional R takes finite values is referred to as the domain of R . The functional R is called proper, if $\mathcal{D}(R) \neq \emptyset$, see Rockafellar (1970).

We will investigate the relation between the following optimization problems:

$$\begin{aligned} \min \{R(\beta): G(\mathbf{X}\beta) \leq q\} & \quad (G\text{-constrained}) & 24. \\ \min \{G(\mathbf{X}\beta) + \gamma R(\beta): \beta \in \mathbb{R}^p\} & \quad (\text{penalized}) & 25. \\ \min \{G(\mathbf{X}\beta): R(\beta) \leq c\} & \quad (R\text{-constrained}), & 26. \end{aligned}$$

for parameters $q, \gamma, c \geq 0$. Here we simplify the notation by $G(\mathbf{X}\beta) \equiv G(\mathbf{X}\beta; Y)$. We will show that Equations 24–26 are equivalent in the sense that any solution β_* of one of these optimization problems also provides a solution of the others.

The proof will use the following saddle-point theorem, which is a well known result from convex analysis.

Lemma 1. Suppose that $\Phi, \Psi: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ are proper convex functionals and denote by $\mathcal{L}: \mathbb{R}^p \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R} \cup \{\infty\}$ the Lagrangian defined by $\mathcal{L}(\beta, \mu) = \Phi(\beta) + \mu\Psi(\beta)$.

1. Let $(\beta_*, \mu_*) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0}$ be a saddle-point of $\mathcal{L}$, that is, for all $(\beta, \mu) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0}$,

$$\mathcal{L}(\beta_*, \mu) \leq \mathcal{L}(\beta_*, \mu_*) \leq \mathcal{L}(\beta, \mu_*).$$

Then $\beta_* \in \operatorname{argmin} \{\Phi(\beta): \Psi(\beta) \leq 0\}$.

2. If, additionally, Slater’s condition

$$\mathcal{D}(\Phi) \cap \{\beta \in \mathbb{R}^p: \Psi(\beta) < 0\} \neq \emptyset$$

holds and $\beta_* \in \operatorname{argmin} \{\Phi(\beta): \Psi(\beta) \leq 0\}$, then there exists a Lagrangian multiplier $\mu_* \geq 0$ such that (β_*, μ_*) is a saddle-point of $\mathcal{L}$.

Proof. Part 1 follows from Theorem 28.3 in Rockafellar (1970).

Part 2 follows from Corollary 28.2.1 and Theorem 28.3 in Rockafellar (1970). $\square$

We next derive the equivalence of the G -constrained and penalized formulation, i.e., of Equations 24 and 25.

Theorem 7 (Equivalence of Equations 24 and 25). Let $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ and $G: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be proper convex functionals and let $\mathbf{X} \in \mathbb{R}^{n \times p}$. Then Equations 24 and 25 are equivalent in the following sense:

1. If β_* is a solution of Equation 25 for some $\gamma > 0$, then β_* solves Equation 24 with $q = G(\mathbf{X}\beta_*)$.
2. For a given $q > 0$, suppose that $G(\mathbf{X}\beta_0) < q$ for some $\beta_0 \in \mathcal{D}(R)$ and that we have

$$\operatorname{argmin}(R) \cap \{\beta: G(\mathbf{X}\beta) \leq q\} = \emptyset. \quad 27.$$

Then, any solution β_* of Equation 24 satisfies $G(\mathbf{X}\beta_*) = q$ and there exists some $\gamma > 0$ such that β_* also solves Equation 25.

In the special case of strictly convex penalties R and $G(\mathbf{X}\beta) = \frac{1}{2} \|Y - \mathbf{X}\beta\|^2$, Theorem 7 seems to be well known but it is hard to find an accessible reference. For an infinite-dimensional setting such a proof can be found in Vasin (1970) and Ivanov et al. (2002). Below we include a short proof for the convex penalties (not necessarily strict convex), but for finite-dimensional ground space. Note that if R is strictly convex and z_0 denotes its unique minimizer, then the condition of Equation 27 reads $G(\mathbf{X}z_0) > q$, which simply means that the unique minimizer of R is not feasible for Equation 24.

Proof of Theorem 7. Part 1: Suppose first that β_* is a solution of Equation 25 for some $\gamma > 0$ and take $q = G(\mathbf{X}\beta_*)$, $\Psi(\beta) = G(\mathbf{X}\beta) - q$. Then we have $\Psi(\beta_*) = 0$ which implies that $R(\beta_*) + \mu\Psi(\beta_*) = R(\beta_*) + \frac{1}{\gamma}\Psi(\beta_*)$ for all $\mu \geq 0$. Further, since β_* is a minimizer of $G(\mathbf{X}\beta; Y) + \gamma R(\beta)$, we also have $\gamma R(\beta_*) + (G(\mathbf{X}\beta_*) - q) \leq \gamma R(\beta) + (G(\mathbf{X}\beta) - q)$, which yields $R(\beta_*) + \frac{1}{\gamma}\Psi(\beta_*) \leq R(\beta) + \frac{1}{\gamma}\Psi(\beta)$ for all $\beta \in \mathbb{R}^p$. In total, we have verified that

$$R(\beta_*) + \mu\Psi(\beta_*) \leq R(\beta_*) + \frac{1}{\gamma}\Psi(\beta_*) \leq R(\beta) + \frac{1}{\gamma}\Psi(\beta) \quad \text{all } (\beta, \mu) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0}.$$

Hence $(\beta_*, 1/\gamma)$ is a saddle-point of the Lagrangian $\mathcal{L}(\beta, \mu) = R(\beta) + \mu\Psi(\beta)$. Lemma 1 implies that $\beta_* \in \operatorname{argmin}\{R(\beta): \Psi(\beta) \leq 0\}$. Recalling that $\Psi(\beta) = G(\mathbf{X}\beta) - q$ shows that β_* is a solution of Equation 24.

Part 2: Now suppose that $\beta_* \in \mathbb{R}^p$ is a solution of the constrained problem in Equation 24 for some $q > 0$. According to our assumption there exists $\beta_0 \in \mathbb{R}^p$ with $G(\mathbf{X}\beta_0) < q$. Consequently, the convex functionals $R(\beta)$ and $\Psi(\beta) = G(\mathbf{X}\beta) - q$ satisfy Slater's condition. According to Lemma 1 the corresponding Lagrangian $\mathcal{L}(\beta, \mu) = R(\beta) + \mu\Psi(\beta)$ admits a saddle-point $(\beta_*, \mu_*) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0}$. In particular, β_* minimises $R(\beta) + \mu_*G(\mathbf{X}\beta)$. It remains to show that $\mu_* > 0$. To that end, suppose to the contrary that $\mu_* = 0$. Then $\beta_* \in \operatorname{argmin} R$. Since β_* is also a solution of Equation 24, it in particular satisfies the constraint in Equation 24, which implies

$$\beta_* \in \operatorname{argmin}(R) \cap \{\beta \in \mathbb{R}^p: G(\mathbf{X}\beta) \leq q\}$$

This contradicts the assumption in Equation 27. Hence $\mu_* > 0$ and further $G(\mathbf{X}\beta_*) = q$. This concludes the proof of the theorem after taking $\gamma = 1/\mu_*$. $\square$

Next we show the equivalence between the penalized and R -constrained formulation, i.e., Equations 25 and 26. Together with Theorem 7 this also implies that the G - and R -constrained formulation are equivalent, too.

Theorem 8 (Equivalence of Equations 25 and 26). Let $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ and $G: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be proper convex functionals and let $\mathbf{X} \in \mathbb{R}^{n \times p}$. Then Equations 25 and 26 are equivalent in the following sense:

1. If β_* is a solution of Equation 25 for some $\gamma > 0$, then β_* solves Equation 26 with $c = R(\beta_*)$.
2. For a given $c > 0$, suppose $R(\beta_0) < c$ and $G(\mathbf{X}\beta_0) < \infty$ for some $\beta_0 \in \mathbb{R}^p$ and let β_* be a solution of Equation 26. Then there exists $\gamma \geq 0$ such that β_* solves Equation 25 and $\gamma(R(\beta_*) - c) = 0$.

Proof. Part 1: Suppose first that β_* is a solution of Equation 25 for some $\gamma > 0$ and take $c = R(\mathbf{X}\beta_*)$, $\Phi(\beta) = G(\mathbf{X}\beta)$, $\Psi(\beta) = R(\beta) - c$. Then we have $\Psi(\beta_*) = 0$ which implies that $\Phi(\beta_*) + \mu\Psi(\beta_*) = \Phi(\beta_*) + \gamma\Psi(\beta_*)$ for all $\mu \geq 0$. Further, since β_* is a minimizer of $G(\mathbf{X}\beta) + \gamma R(\beta)$, we also have $G(\mathbf{X}\beta_*) + \gamma(R(\beta_*) - c) \leq G(\mathbf{X}\beta) + \gamma(R(\beta) - c)$ for $\beta \in \mathbb{R}^p$, which yields $\Phi(\beta_*) + \gamma\Psi(\beta_*) \leq \Phi(\beta) + \gamma\Psi(\beta)$. In total, we have verified the inequalities

$$\Phi(\beta_*) + \mu\Psi(\beta_*) \leq \Phi(\beta_*) + \gamma\Psi(\beta_*) \leq \Phi(\beta) + \gamma\Psi(\beta) \quad \text{for } (\beta, \mu) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0}.$$

Hence (β_*, γ) is a saddle-point of the corresponding Lagrangian $\mathcal{L}(\beta, \mu) = \Phi(\beta) + \mu\Psi(\beta)$. Lemma 1 therefore implies that $\beta_* \in \operatorname{argmin}\{\Phi(\beta) : \Psi(\beta) \leq 0\}$. Recalling that $\Phi(\beta) = G(\mathbf{X}\beta)$, $\Psi(\beta) = R(\beta) - c$ shows that β_* is a solution of Equation 24.

Part 2: Suppose that $\beta_* \in \mathbb{R}^p$ is a solution of Equation 26 of some $c > 0$. According to our assumption there exists $\beta_0 \in \mathbb{R}^p$ with $R(\beta_0) < q$. Consequently, the convex functionals $\Phi(\beta) = G(\mathbf{X}\beta)$ and $\Psi(\beta) = R(\beta) - c$ satisfy Slater's condition. According to Lemma 1 the corresponding Lagrangian $\mathcal{L}(\beta, \mu) = G(\mathbf{X}\beta) + \mu(R(\beta) - c)$ admits a saddle-point $(\beta_*, \gamma) \in \mathbb{R}^p \times \mathbb{R}_{\geq 0}$. In particular, β_* minimises $G(\mathbf{X}\beta) + \gamma R(\beta)$. Since β_* is feasible, we have $R(\beta_*) - c \leq 0$. Further, for all $\mu \geq 0$,

$$G(\mathbf{X}\beta_*) + \mu(R(\beta_*) - c) \leq G(\mathbf{X}\beta_*) + \gamma(R(\beta_*) - c),$$

which implies $\gamma(R(\beta_*) - c) = 0$. □

We emphasize again that the relation between the threshold q and the penalty γ in Theorem 7 is only given in an implicit manner. Therefore, the unconstrained problem in Equation 25 cannot be immediately used for solving the (more difficult) constrained problem in Equation 24. The same statement applies to the equivalence of the penalized and R -constrained formulation. As shown in the next subsection, a more explicit relation between the G -constrained formulation and yet another different unconstrained optimization problem can be derived from the Fenchel duality.

The derived equivalences in particular apply to the MIND in Equation 9 by taking

$$G(v) = \max_{a \in \mathcal{A}} \left\{ \frac{\|\Pi_a(v - Y)\|_2}{w_a} - s_a \right\}$$

for $v \in \mathbb{R}^n$, where Π_a are given probe functionals, w_a and s_a certain weights and Y the given data, see Definition 1. Note that the functional G is obviously convex and proper.

A.2. Fenchel Duality

For a proper convex functional $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$, we denote by

$$\begin{aligned} R^*: \mathbb{R}^p &\rightarrow \mathbb{R} \cup \{\infty\} \\ \mu &\mapsto \sup \{ \langle \mu, \beta \rangle - R(\beta) : \beta \in \mathbb{R}^p \} \end{aligned}$$

the Fenchel conjugate of R , and by $\partial R(\beta_0)$ the subdifferential of R at β_0 , i.e., $\beta^* \in \partial R(\beta_0)$ if and only if

$$R(\beta_0) < \infty \quad \text{and} \quad \langle \beta - \beta_0, \beta^* \rangle + R(\beta_0) \leq R(\beta) \quad \text{for all } \beta \in \mathbb{R}^p,$$

see Rockafellar (1970).

Definition 4. Let $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ and $G: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ be two proper convex functionals and let $\mathbf{X} \in \mathbb{R}^{n \times p}$. Then,

$$\min_{\beta \in \mathbb{R}^p} G(\mathbf{X}\beta) + R(\beta) \tag{primal} \tag{28}$$

$$\min_{\mu \in \mathbb{R}^n} G^*(\mu) + R^*(-\mathbf{X}^T \mu) \tag{dual} \tag{29}$$

are referred to as the primal and dual minimization problem, respectively, corresponding to R , G and $\mathbf{X}$.

Theorem 9 (Fenchel's duality theorem). Suppose that $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ and $G: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ are proper convex functionals, and let $\mathbf{X} \in \mathbb{R}^{n \times p}$. Suppose further that there exists $\beta_0 \in \mathbb{R}^p$ such that $R(\beta_0) < \infty$ and that G is continuous at $\mathbf{X}\beta_0$. Then, for any $(\beta_*, \mu_*) \in \mathbb{R}^p \times \mathbb{R}^n$, the following statements are equivalent:

1. β_* is a solution of the primal problem in Equation 28 and μ_* a solution of the dual problem in Equation 29.
2. The Kuhn-Tucker conditions are satisfied, i.e.,

$$\mu_* \in \partial G(\mathbf{X}\beta_*) \tag{30}$$

$$-\mathbf{X}^T \mu_* \in \partial R(\beta_*). \tag{31}$$

Proof. See (Rockafellar 1970, Section 31). □

If one takes $G(v) = \frac{1}{2} \|Y - v\|_2^2$, then Theorem 9 yields the following result:

Theorem 10. Let $R: \mathbb{R}^p \rightarrow \mathbb{R} \cup \{\infty\}$ be a proper convex functional and let $Y \in \mathbb{R}^n$. Then,

$$\operatorname{argmin}_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|Y - \mathbf{X}\beta\|_2^2 + R(\beta) \right\} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{2} \|\mathbf{X}\beta\|_2^2 + R^*(\mathbf{X}^T(Y - \mathbf{X}\beta)) \right\}.$$

Proof. If β_* is a minimizer of $\frac{1}{2} \|Y - \mathbf{X}\beta\|_2^2 + R(\beta)$, then it is a solution of the primal problem in Equation 28 corresponding to $G(v) = \frac{1}{2} \|Y - v\|_2^2$. In this case, one easily computes $G^*(\mu) = \frac{1}{2} \|\mu\|_2^2 + \langle Y, \mu \rangle$, see Rockafellar (1970). Now let $\mu_* \in \mathbb{R}^n$ denote a solution of the dual problem in Equation 29. Since R is proper there exists $\beta_0 \in \mathbb{R}^p$ such

that $R(\beta_0) < \infty$ and obviously G is continuous at $\mathbf{X}\beta_0$. Hence we can apply Theorem 9 which implies that the Kuhn-Tucker conditions in Equations 30 and 31 are satisfied. In particular, Equation 30 implies $\mu_\star = \mathbf{X}\beta_\star - Y$, and therefore

$$\begin{aligned} G^\star(\mu_\star) &= G^\star(\mathbf{X}\beta_\star - Y) \\ &= \frac{1}{2} \|\mathbf{X}\beta_\star - Y\|_2^2 + \langle Y, \mathbf{X}\beta_\star - Y \rangle \\ &= \frac{1}{2} \|\mathbf{X}\beta_\star\|_2^2 - \frac{1}{2} \|Y\|_2^2. \end{aligned}$$

Because $\mu_\star$ is a solution of the dual problem in Equation 29, this shows that $\beta_\star$ minimizes $\frac{1}{2} \|\mathbf{X}\beta\|_2^2 + R^\star(\mathbf{X}^\top(Y - \mathbf{X}\beta))$. Similar arguments show that the converse relation also holds. $\square$

B. PROOFS

Here we provide proofs for the assertions in the article for the sake of completeness, noting that many of them are known but scattered in the literature. These assertions can be seen as special cases of the results obtained in Appendix A. But we favor simple and direct proofs whenever it is possible.

B.1. Proofs in Section 3

In this subsection, we give proofs for the assertions in Section 3 in the article. Note that the following proofs do not rely on any results from Appendix A.

Proof of Theorem 1. Because $(\phi_\lambda)_{\lambda \in \Lambda}$ is an orthonormal basis, we can uniquely write any element in $\mathbb{R}^n$ as linear combination $\beta = \sum x_\lambda \phi_\lambda$ with coefficients $x_\lambda = \langle \phi_\lambda, \beta \rangle$. Hence $\hat{\beta} = \sum \hat{x}_\lambda \phi_\lambda$ is a solution of the stated optimization problem if and only if every coefficient $\hat{x}_\lambda$ is a solution of the one-dimensional optimization problem

$$\begin{cases} \min_{x \in \mathbb{R}} & |x_\lambda|^r \\ \text{s. t.} & |\langle \phi_\lambda, Y \rangle - x_\lambda| \leq q_\lambda. \end{cases}$$

The unique minimizer is given by the soft thresholding $\hat{x}_\lambda = \eta^{(\text{soft})}(\langle \phi_\lambda, Y \rangle, q_\lambda)$, which yields the desired characterization using the objective $\sum_{\lambda \in \Lambda} |\langle \phi_\lambda, \beta \rangle|^r$ for any $r > 0$.

In the special case $r = 2$, we have $\sum_{\lambda \in \Lambda} |\langle \phi_\lambda, \beta \rangle|^2 = \sum_{i=1}^n \beta_i^2$, which shows the second claim. $\square$

Proof of Theorem 2. This immediately follows from the variational characterization of soft thresholding given in Theorem 1 and the relation

$$\eta^{(\text{JS})}(x, q) = \eta^{(\text{soft})}\left(x, \frac{q^2}{|x|}\right) \text{ for } x \neq 0.$$

$\square$

Proof of Theorem 3. The proof is elementary and similar to the one of Theorem 1. $\square$

Proof of Theorem 4. By expanding any function $f \in L^2(\Omega)$ in a wavelet series $f = \sum_{\lambda \in \Lambda} x_\lambda \phi_\lambda$, with uniquely determined coefficients $x_\lambda \in \mathbb{R}$, the stated minimization is equivalent to

$$\begin{cases} \min_x & \sum_{\lambda \in \Lambda} |x_\lambda|^r \\ \text{s. t.} & \max_{\lambda \in \Lambda} \frac{|\kappa_\lambda^{-1} \langle v_\lambda, g \rangle - x_\lambda|}{q_\lambda} \leq 1. \end{cases}$$

The solution of the latter optimization problem is given by component-wise soft thresholding of $\kappa_\lambda^{-1} \langle v_\lambda, g \rangle$ with thresholds q_λ (see the proof of Theorem 1). This results in $\hat{f}^{(\text{WV})}$. $\square$

B.2. Proofs in Section 4

In this subsection, we provide proofs for the assertions in Section 4 in the article. They rely on the results from Appendix A.

To establish Theorem 5, we apply Theorem 10 to the case where R is the block ℓ^1 -penalty. To that end we first compute its Fenchel conjugate. Recall that subsets Λ_a are assumed to be disjoint.

Lemma 2. The Fenchel conjugate of the functional $R(\beta) = \sum_{a \in \mathcal{A}} w_a \|\Phi_a \beta\|_2$ is given by

$$R^*(\mu) = \begin{cases} 0 & \text{if } \max_{a \in \mathcal{A}} \frac{\|\Phi_a \mu\|_2}{w_a} \leq 1 \\ \infty & \text{otherwise.} \end{cases}$$

Proof. This can be verified by straightforward computation. $\square$

We are now ready to prove Theorem 5.

Proof of Theorem 5. Theorem 10 implies that $\beta_\star \in \mathbb{R}^p$ is a minimizer of the adaptive group lasso in Equation 19 if and only if it is a minimizer of the functional $\frac{1}{2} \|\mathbf{X}\beta\|_2^2 + R^*(\mathbf{X}^\text{T}(Y - \mathbf{X}\beta))$ with the particular choice $R(\beta) = \gamma \sum_{a \in \mathcal{A}} w_a \|\Phi_a \beta\|_2$. According to Lemma 2 this is equivalent to the fact that $\beta_\star$ minimizes $\frac{1}{2} \|\mathbf{X}\beta\|_2^2$ under the constraint

$$\max_{a \in \mathcal{A}} \frac{\|\Phi_a \mathbf{X}^\text{T}(Y - \mathbf{X}\beta)\|_2}{w_a} \leq \gamma.$$

This means that $\beta_\star$ is the unique minimizer of Equation 20 as we intended to show. $\square$

Proof of Corollary 1. By specializing $(\phi_\lambda)_{\lambda \in \Lambda}$ to the standard basis in $\mathbb{R}^n$, letting the groups consist of single elements, and taking all weights equal to one, the constraint in Theorem 5 reduces to $\|\mathbf{X}^\text{T}(Y - \mathbf{X}\beta)\|_\infty \leq \gamma$. Further, in such a situation the adaptive group lasso reduces to the standard lasso. Hence the claim follows from Theorem 5. $\square$

Proof of Corollary 2. According to Theorem 3, the estimators in Equations 15 and 20 coincide. According to Theorem 5, the estimators in Equations 20 and 19 coincide. $\square$

Proof of Theorem 6. It follows immediately from Theorems 7 and 8. $\square$

LITERATURE CITED

- Ivanov VK, Vasin VV, Tanana VP. 2002. *Theory of Linear Ill-Posed Problems and Its Applications*. Utrecht, Neth.: De Gruyter. 2nd ed.
- Rockafellar RT. 1970. *Convex Analysis*. Princeton, NJ: Princeton Univ. Press
- Vasin VV. 1970. The connection between certain variational methods for the approximate solution of ill posed problems. *Mat. Zametki* 7:265–72