Supplementary Material

Differentiating Minimal-Norm Solutions
to Parametric Optimization Problems

This document supplements the main paper. Sections, equations, figures, tables, and results in this document use an "SM" prefix to distinguish them from those in the main paper.

SM1 Failure of Tikhonov regularization

Remark SM1 (Failure of **(R')** in Figure SM1)**.** Consider the problem $\max_{\|x\|_\infty \le 1} \langle \theta, x \rangle$, at $\theta_0 = (0, -1)$. The Tikhonov regularized solution reads, by separability,

$$(x^\star_\varepsilon(\theta))_i = \max(\min(+1, \theta_i/\varepsilon), -1) = \Pi_{[-1,+1]}(\theta_i/\varepsilon),$$

$$J = \operatorname{diag}(d_1, d_2) \in J_{x^\star_\varepsilon}(\theta), \qquad d_i \in \begin{cases} \{0\} & |\theta_i| > \varepsilon, \\ \{1/\varepsilon\} & |\theta_i| < \varepsilon, \\ [0, 1/\varepsilon] & |\theta_i| = \varepsilon. \end{cases}$$

At $\theta_0 = (0, -1)$ and $0 < \varepsilon < 1$, every such Jacobian has $d_1 = 1/\varepsilon$ and $d_2 = 0$. The first entry diverges as $\varepsilon \downarrow 0$, consistently with the discontinuity of the minimal-norm solution. In the composite formulation of Section 4 of the main paper, take $f(x, \theta) = -\langle \theta, x \rangle$ and $g = \iota_{[-1,1]^2}$. The corresponding blocks satisfy $H^f = 0$, $M^f = -I$, $P = 0$, and $Q = \operatorname{diag}(1, 0)$. Thus $\Sigma = 0$ and $V = -\gamma Q \neq 0$, so $\operatorname{range}(V) \not\subseteq \operatorname{range}(\Sigma)$: the range condition **(R')** fails.

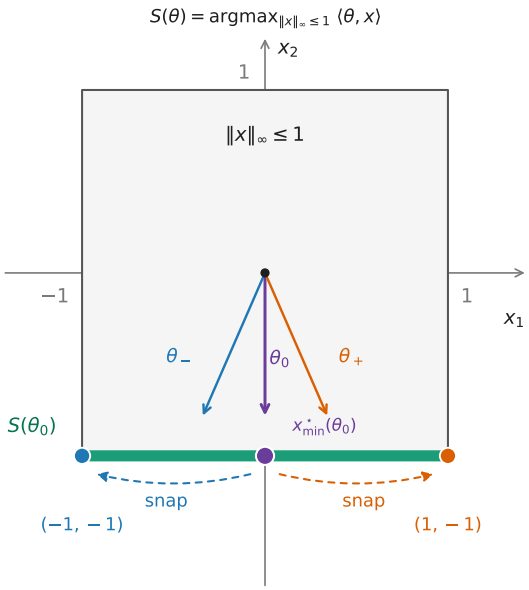


Figure SM1: A simple composite problem, $\max_{\|x\|_\infty \le 1} \langle \theta, x \rangle$. The minimal-norm solution is $x^\star_{\min}(\theta) = \operatorname{sign}(\theta)$, with $\operatorname{sign}(0) = 0$, which is not continuous at the represented $\theta_0 = (0, -1)$ and hence is not path-differentiable; the range conditions must therefore fail (see Remark SM1). Here $\mathcal{S}(\theta_0)$, shown in green, is an edge of the ball and $x^\star_{\min}(\theta_0) = (0, -1)$. Perturbing θ_0 toward θ_- or θ_+ can collapse the solution set to a singleton and make the minimal-norm solution jump to a corner.

SM2 Details on NYC 311 Huber Data Poisoning

This section provides additional details for Sections 5.1 and A.1 of the main paper. Let $A \in \mathbb{R}^{n_{\mathrm{tr}} \times p}$ and $\theta_0 \in \mathbb{R}^{n_{\mathrm{tr}}}$ denote the lower-level design and clean response, and let A_{val} and b_{val} denote the validation data. For a perturbation budget $\tau > 0$, the bilevel problem is

$$\begin{aligned} &\underbrace{\max_{\|\theta-\theta_0\|_2 \le \tau} \frac{1}{2n_{\mathrm{val}}} \|A_{\mathrm{val}} x^\star(\theta) - b_{\mathrm{val}}\|_2^2}_{\text{upper-level problem}} \\ &\text{subject to} \quad x^\star(\theta) \in \underbrace{\operatorname*{argmin}_{x \in \mathbb{R}^p} \sum_{i=1}^{n_{\mathrm{tr}}} h_\delta\big((Ax-\theta)_i\big)}_{\text{lower-level problem}}. \end{aligned} \tag{SM1}$$

Here, the scalar Huber loss with parameter $\delta > 0$ is

$$h_\delta(r) = \begin{cases} \frac{1}{2} r^2, & |r| \le \delta, \\ \delta |r| - \frac{1}{2}\delta^2, & |r| > \delta. \end{cases} \tag{SM2}$$

For vector arguments, we write $H_\delta(r) := \sum_{i=1}^{n_{\mathrm{tr}}} h_\delta(r_i)$, so the lower-level objective is $H_\delta(Ax - \theta)$.

Proposition SM2. *Let $\delta > 0$. The summed Huber loss satisfies*

$$H_\delta = \delta \left((\|\cdot\|_1)^* + \frac{\delta}{2} \|\cdot\|_2^2 \right)^*. \tag{SM3}$$

The corresponding scalar identity holds in dimension one. An equivalent minimization form of the dual of the lower-level problem in (SM1) *is the strongly convex problem*

$$\begin{aligned} u^\star(\theta) = \operatorname*{argmin}_{u \in \mathbb{R}^{n_{\mathrm{tr}}}} & \frac{1}{2}\|u + \theta\|_2^2 \\ \textit{subject to} \quad & A^\top u = 0, \qquad \|u\|_\infty \le \delta. \end{aligned} \tag{SM4}$$

If $p(\theta) := \min_x H_\delta(Ax - \theta)$, their optimal values satisfy

$$p(\theta) = \frac{1}{2}\|\theta\|_2^2 - \frac{1}{2}\|u^\star(\theta) + \theta\|_2^2. \tag{SM5}$$

Define $I_0 = \{i : |u_i^\star| < \delta\}$, $I_+ = \{i : u_i^\star = \delta\}$, and $I_- = \{i : u_i^\star = -\delta\}$. The solution set of the primal lower-level problem is

$$\mathcal{S}(\theta) = \left\{ x \in \mathbb{R}^p : \begin{array}{r} A_{I_0,:} x = \theta_{I_0} + u^\star_{I_0}, \\ A_{I_+,:} x \ge \theta_{I_+} + \delta \mathbf{1}, \\ A_{I_-,:} x \le \theta_{I_-} - \delta \mathbf{1} \end{array} \right\}. \tag{SM6}$$

Proof. First, we justify (SM3). Let ι_C be the indicator function of a set C, equal to zero on C and to $+\infty$ elsewhere. Using the definition of the Fenchel conjugate (Bauschke and Combettes, 2017), we obtain

$$\begin{aligned} (\|\cdot\|_1)^*(y) &= \sup_x \{\langle x, y\rangle - \|x\|_1\} = \iota_{B_{\ell_\infty}(0,1)}(y), \\ g_\delta(y) &:= \left((\|\cdot\|_1)^* + \frac{\delta}{2}\|\cdot\|_2^2 \right)(y) \\ &= \iota_{B_{\ell_\infty}(0,1)}(y) + \frac{\delta}{2}\|y\|_2^2. \end{aligned}$$

Consequently,

$$\begin{aligned} g_\delta^*(r) &= \sup_y \left\{ \langle y, r\rangle - \iota_{B_{\ell_\infty}(0,1)}(y) - \frac{\delta}{2}\|y\|_2^2 \right\} \\ &= \sup_{\|y\|_\infty \le 1} \left\{ \langle y, r\rangle - \frac{\delta}{2}\|y\|_2^2 \right\} \\ &= \sum_i \sup_{|y_i|\le 1} \left\{ y_i r_i - \frac{\delta}{2} y_i^2 \right\} \\ &= \sum_i \begin{cases} r_i - \delta/2, & r_i > \delta, \\ r_i^2/(2\delta), & r_i \in [-\delta, \delta], \\ -r_i - \delta/2, & r_i < -\delta. \end{cases} \end{aligned}$$

Multiplication by δ yields the sum of the scalar expressions in (SM2), proving (SM3).

We next derive the dual at the level of optimal values. Since H_δ is finite and continuous, strong duality applies (Bauschke and Combettes, 2017, Theorem 15.23 and Proposition 15.24). Thus

$$\begin{aligned} p(\theta) &= \min_x H_\delta(Ax - \theta) \\ &= \delta \min_x \sup_{\|y\|_\infty \le 1} \left\{ \langle y, Ax - \theta\rangle - \frac{\delta}{2}\|y\|_2^2 \right\} \\ &= \delta \max_{\|y\|_\infty \le 1} \left\{ -\langle y, \theta\rangle - \frac{\delta}{2}\|y\|_2^2 + \inf_x \langle A^\top y, x\rangle \right\} \\ &= \delta \max_{\substack{\|y\|_\infty \le 1 \\ A^\top y = 0}} \left\{ -\langle y, \theta\rangle - \frac{\delta}{2}\|y\|_2^2 \right\} \\ &= \max_{\substack{\|u\|_\infty \le \delta \\ A^\top u = 0}} \left\{ -\langle u, \theta\rangle - \frac{1}{2}\|u\|_2^2 \right\} \qquad (u = \delta y) \\ &= \frac{1}{2}\|\theta\|_2^2 - \min_{\substack{\|u\|_\infty \le \delta \\ A^\top u = 0}} \frac{1}{2}\|u + \theta\|_2^2. \end{aligned}$$

Here $\inf_x \langle A^\top y, x\rangle$ equals zero when $A^\top y = 0$ and $-\infty$ otherwise. Completing the square gives the last equality. The primal minimum equals the maximum of the concave dual objective; it is the constant $\|\theta\|_2^2/2$ minus the minimum of the shifted projection objective, as stated in (SM5). The feasible set in (SM4) is a nonempty polytope, being the intersection of a linear subspace with a bounded polyhedral ball. Its objective is strongly convex, so the minimizer $u^\star$ exists and is unique.

We now establish (SM6). The point $u = 0$ satisfies the equality constraint and strictly satisfies both box inequalities because $\delta > 0$. Hence the constraints satisfy Slater's condition, and the KKT conditions are necessary and sufficient. Let x be the multiplier associated with $-A^\top u = 0$, and let $\mu_+ \ge 0$ and $\mu_- \ge 0$ be the multipliers associated with $u - \delta\mathbf{1} \le 0$ and $-u - \delta\mathbf{1} \le 0$, respectively. The KKT system is

$$\begin{cases} u + \theta - Ax + \mu_+ - \mu_- = 0, \\ A^\top u = 0, \\ 0 \le \mu_+ \ \perp \ (u - \delta\mathbf{1}) \le 0, \\ 0 \le \mu_- \ \perp \ (-u - \delta\mathbf{1}) \le 0, \end{cases}$$

where $\perp$ denotes orthogonality. At $u = u^\star$, splitting the indices as in the statement gives

$$\begin{aligned}
\text{on } I_+ : \quad & (\mu_-)_{I_+} = 0, \\
& \delta\mathbf{1} + \theta_{I_+} - A_{I_+,:}x + (\mu_+)_{I_+} = 0, \\
\text{on } I_0 : \quad & (\mu_-)_{I_0} = (\mu_+)_{I_0} = 0, \\
& u^\star_{I_0} + \theta_{I_0} - A_{I_0,:}x = 0, \\
\text{on } I_- : \quad & (\mu_+)_{I_-} = 0, \\
& -\delta\mathbf{1} + \theta_{I_-} - A_{I_-,:}x - (\mu_-)_{I_-} = 0.
\end{aligned}$$

Using the nonnegativity of the multipliers, the boundary equations are equivalent to

$$\begin{aligned}
A_{I_+,:}x = \theta_{I_+} + \delta\mathbf{1} + (\mu_+)_{I_+} \quad & \text{for some } (\mu_+)_{I_+} \geq 0 \\
\iff \quad & A_{I_+,:}x \geq \theta_{I_+} + \delta\mathbf{1}, \\
A_{I_-,:}x = \theta_{I_-} - \delta\mathbf{1} - (\mu_-)_{I_-} \quad & \text{for some } (\mu_-)_{I_-} \geq 0 \\
\iff \quad & A_{I_-,:}x \leq \theta_{I_-} - \delta\mathbf{1}.
\end{aligned}$$

Together with the equality on I_0, these are precisely the conditions in (SM6). Conversely, any x satisfying those conditions yields nonnegative multipliers satisfying the KKT system. Equivalently, these conditions say that $u_i^\star = h'_\delta((Ax-\theta)_i)$ for every i; combined with $A^\top u^\star = 0$, they characterize the minimizers of the differentiable convex function $x \mapsto H_\delta(Ax-\theta)$, which concludes the proof. $\square$

SM3 Details on the Illustration of the Different Regimes

We present the reasoning behind the experiment in Section 5.3 of the main paper. Here $\theta \in \mathbb{R}^n_+$ is the observed vector of singular values, ordered so that $\theta_1 \geq \cdots \geq \theta_n$; the experiment uses $n = 40$. The optimization variables are vectors, and differentiation is with respect to θ, not through the singular-value decomposition. The penalty $\alpha > 0$ is held fixed when taking these derivatives.

We recall the four optimization problems, with $h(m) = \frac{1}{2}\|m-\theta\|_2^2 + \alpha\|m\|_1$:

$$\min_{m\in\mathbb{R}^n} h(m), \tag{SMP1}$$

$$\min_{x_1,x_2\in\mathbb{R}^n} \underbrace{\tfrac{1}{2}\|x_1+x_2-\theta\|_2^2 + \alpha(\|x_1\|_1 + \|x_2\|_1)}_{\widetilde{h}(x_1,x_2)}, \tag{SMP2}$$

$$\min_{m\in\mathbb{R}^n} v(m), \qquad v(m) := \min_{x_1+x_2=m} \left\{ \tfrac{1}{2}\|m-\theta\|_2^2 + \alpha(\|x_1\|_1 + \|x_2\|_1) \right\}, \tag{SMP3}$$

$$\min_m \min_{x_1+x_2=m} \widetilde{h}_\varepsilon(x_1,x_2), \qquad \widetilde{h}_\varepsilon(x_1,x_2) := \widetilde{h}(x_1,x_2) + \tfrac{\varepsilon}{2}(\|x_1\|_2^2 + \|x_2\|_2^2). \tag{SMP4}$$

Proposition SM3 (Solution sets and their regularization)**.** *Let $m^\star = \operatorname{prox}_{\alpha\|\cdot\|_1}(\theta)$ be the unique minimizer of* (SMP1)*. Then $v = h$, the problems* (SMP1)–(SMP3) *have optimal value $h(m^\star)$, and the outer minimizer of* (SMP3) *is $m^\star$. The solution set of* (SMP2)*, equivalently the inner solution set of* (SMP3) *at $m^\star$, is*

$$\mathcal{S}(m^\star) = \{(x_1,x_2) : x_1 + x_2 = m^\star, \quad \|x_1\|_1 + \|x_2\|_1 = \|m^\star\|_1\}.$$

This set is nonempty, convex and compact. It contains $\{(tm^\star, (1-t)m^\star) : t \in [0,1]\}$, and its Euclidean minimal-norm element is $(m^\star, m^\star)/2$. For every $\varepsilon > 0$, the unique minimizer of (SMP4) *is $(m^\star, m^\star)/(2+\varepsilon)$.*

Proof. For every m and every pair with $x_1 + x_2 = m$, the triangle inequality gives $\|x_1\|_1 + \|x_2\|_1 \geq \|m\|_1$. Equality is attained by $(x_1, x_2) = (m, m)/2$, so $v(m) = h(m)$ for every m. Thus $m^\star$ is the outer minimizer of (SMP3), and its inner solution set is $\mathcal{S}(m^\star)$.

For any pair (x_1, x_2), $\widetilde{h}(x_1, x_2) \geq h(x_1 + x_2) \geq h(m^\star)$, and equality is attained by $(m^\star, m^\star)/2$. This proves $\min \widetilde{h} = h(m^\star)$ and the asserted description of its minimizers. As the solution set of a continuous convex problem, $\mathcal{S}(m^\star)$ is closed and convex. It contains the stated segment because

$$tm^\star + (1-t)m^\star = m^\star, \qquad \|tm^\star\|_1 + \|(1-t)m^\star\|_1 = \|m^\star\|_1 \quad (0 \leq t \leq 1).$$

It is bounded, hence compact, since $\|x_j\|_2 \leq \|x_j\|_1 \leq \|m^\star\|_1$ on this set. The first inequality follows from $\sum_i x_{j,i}^2 \leq (\sum_i |x_{j,i}|)^2$.

For any pair with $x_1 + x_2 = m$, the parallelogram identity gives

$$\begin{aligned} \tfrac{1}{2}(\|x_1\|_2^2 + \|x_2\|_2^2) &= \tfrac{1}{4}(\|x_1 - x_2\|_2^2 + \|x_1 + x_2\|_2^2) \\ &= \tfrac{1}{4}(\|m\|_2^2 + \|x_1 - x_2\|_2^2). \end{aligned} \tag{SM7}$$

At $m = m^\star$, the unique minimal-norm pair therefore has $x_1 = x_2 = m^\star/2$.

Finally, for (SMP4), both the triangle inequality and (SM7) attain their lower bounds at $x_1 = x_2 = m/2$. Its outer objective is consequently

$$\begin{aligned} &\tfrac{1}{2}\|m - \theta\|_2^2 + \alpha\|m\|_1 + \tfrac{\varepsilon}{4}\|m\|_2^2 \\ &\quad = \alpha\|m\|_1 + \tfrac{1}{2}\left(\|\theta\|_2^2 - 2\langle m, \theta\rangle + \tfrac{2+\varepsilon}{2}\|m\|_2^2\right). \end{aligned}$$

Put $t = 2/(2+\varepsilon)$. Multiplying this objective by t and adding a constant independent of m gives

$$t\alpha\|m\|_1 + \tfrac{1}{2}\left(t^2\|\theta\|_2^2 - 2\langle m, t\theta\rangle + \|m\|_2^2\right) = t\alpha\|m\|_1 + \tfrac{1}{2}\|m - t\theta\|_2^2.$$

Its unique minimizer is $m_\varepsilon = \operatorname{prox}_{t\alpha\|\cdot\|_1}(t\theta)$. The soft-thresholding formula

$$\operatorname{prox}_{a\|\cdot\|_1}(u) = \operatorname{sign}(u)(|u| - a)_+$$

implies $\operatorname{prox}_{t\alpha\|\cdot\|_1}(tu) = t\operatorname{prox}_{\alpha\|\cdot\|_1}(u)$ for $t > 0$. Hence $m_\varepsilon = 2m^\star/(2+\varepsilon)$, and each half is $m^\star/(2+\varepsilon)$, as claimed. □

The convergence to the minimal-norm solution, established generally by Tikhonov regularization in the main paper (Theorem 36), is explicit here because $1/(2+\varepsilon) \to 1/2$.

A coordinate that stays on the threshold. For $j \in \{1, 2\}$, the forward–backward fixed-point equations are

$$\begin{aligned} z_{j,\varepsilon} &= x_j - \gamma(x_1 + x_2 - \theta + \varepsilon x_j), \\ (x_1, x_2) &= \left(\operatorname{prox}_{\gamma\alpha\|\cdot\|_1}(z_{1,\varepsilon}), \operatorname{prox}_{\gamma\alpha\|\cdot\|_1}(z_{2,\varepsilon})\right). \end{aligned} \tag{SM8}$$

At the regularized solution, both proximal inputs are the vector $(\zeta_i)_{i=1}^n$, where

$$\zeta_i = \frac{(\theta_i - \alpha)_+}{2+\varepsilon} + \gamma\big(\theta_i - (\theta_i - \alpha)_+\big).$$

To verify this, separability and symmetry reduce the calculation to

$$\begin{aligned} z_i &= x_i - \gamma\big((2+\varepsilon)x_i - \theta_i\big), \\ x_i &= \operatorname{prox}_{\gamma\alpha|\cdot|}(z_i) = \begin{cases} z_i - \gamma\alpha, & z_i > \gamma\alpha, \\ 0, & |z_i| \leq \gamma\alpha, \\ z_i + \gamma\alpha, & z_i < -\gamma\alpha. \end{cases} \end{aligned} \tag{SM9}$$

Since $\theta_i \geq 0$, we have $\zeta_i \geq 0$. There are two cases:

	$z_i = \zeta_i$	x_i
$\theta_i > \alpha$	$\dfrac{\theta_i - \alpha}{2+\varepsilon} + \gamma\alpha > \gamma\alpha$	$\dfrac{\theta_i - \alpha}{2+\varepsilon}$
$\theta_i \leq \alpha$	$\gamma\theta_i \leq \gamma\alpha$	$0.$

Substitution into the first equation of (SM9) gives

$$\theta_i > \alpha: \quad \frac{\theta_i - \alpha}{2+\varepsilon} - \gamma\left((2+\varepsilon)\frac{\theta_i - \alpha}{2+\varepsilon} - \theta_i\right) = \frac{\theta_i - \alpha}{2+\varepsilon} + \gamma\alpha,$$
$$\theta_i \leq \alpha: \quad 0 - \gamma\big((2+\varepsilon)0 - \theta_i\big) = \gamma\theta_i.$$

Thus both equations hold. Choose a positive coordinate $\theta_k > 0$ and set $\alpha = \theta_k$ at this base point. Then $\zeta_k = \gamma\alpha$ for every $\varepsilon > 0$, so the chosen coordinate always lies at a kink of soft-thresholding. If $\theta = 0$, every solution is zero and there is no positive coordinate to use for this construction.

Jacobian selections and the three regimes. For $Q_j \in \partial \operatorname{prox}_{\gamma\alpha\|\cdot\|_1}(z_{j,\varepsilon})$, set $Q = \operatorname{diag}(Q_1, Q_2)$. The parameter block of the proximal map is $P = 0$ because α is fixed. The derivative blocks are

$$H^f = \begin{pmatrix} I & I \\ I & I \end{pmatrix}, \qquad M^f = \begin{pmatrix} -I \\ -I \end{pmatrix}, \qquad M_\varepsilon = -\gamma \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix},$$
$$H_\varepsilon = I - Q\big((1-\gamma\varepsilon)I - \gamma H^f\big) \tag{SM10}$$
$$= \begin{pmatrix} I - (1-\gamma-\gamma\varepsilon)Q_1 & \gamma Q_1 \\ \gamma Q_2 & I - (1-\gamma-\gamma\varepsilon)Q_2 \end{pmatrix}.$$

The selected implicit Jacobian is $J_\varepsilon = -H_\varepsilon^{-1} M_\varepsilon$. The diagonal entries of a soft-thresholding Clarke Jacobian are 1 above the threshold, 0 below it, and any number in $[0, 1]$ at equality.

For $i < k$ with $\theta_i = \alpha + \eta$, $\eta > 0$, the proximal input satisfies

$$\begin{aligned} \zeta_i &= \frac{(\alpha + \eta - \alpha)_+}{2+\varepsilon} + \gamma\big(\alpha + \eta - (\alpha + \eta - \alpha)_+\big) \\ &= \frac{\eta}{2+\varepsilon} + \gamma\alpha > \gamma\alpha. \end{aligned}$$

Hence $(Q_1)_{ii} = (Q_2)_{ii} = 1$. At every additional tied coordinate $i \neq k$ with $\theta_i = \alpha$, we also choose both diagonal entries to be 1, whether $i < k$ or $i > k$. At the designated kink, choose

$$(Q_1)_{kk} = 1, \qquad (Q_2)_{kk} = \frac{1}{1 + \gamma\kappa\varepsilon^\delta}, \qquad \kappa > 0, \quad \delta > 0.$$

For $i > k$ with $\theta_i = \alpha - \eta$, $\eta > 0$,

$$\begin{aligned} \zeta_i &= \frac{(\alpha - \eta - \alpha)_+}{2+\varepsilon} + \gamma\big(\alpha - \eta - (\alpha - \eta - \alpha)_+\big) \\ &= \gamma(\alpha - \eta) < \gamma\alpha, \end{aligned}$$

so both diagonal entries are 0. Only coordinate k has differing selections between the two copies. We will obtain the slow, critical and fast regimes for $\delta < 1$, $\delta = 1$ and $\delta > 1$, respectively.

To display all matrix calculations at a readable size, define

$$\mathcal{A} = \{i \neq k : \theta_i \geq \alpha\}, \qquad \mathcal{I} = \{i : \theta_i < \alpha\}, \qquad \mathcal{D}(a, b, c) = a\Pi_{\mathcal{A}} + be_k e_k^\top + c\Pi_{\mathcal{I}},$$

where $\Pi_{\mathcal{A}}$ and $\Pi_{\mathcal{I}}$ are diagonal coordinate projectors. Thus $\mathcal{D}(a,b,c)$ has entries a on $\mathcal{A}$, b at k, and c on $\mathcal{I}$. If the threshold is isolated, these sets are $\{1,\dots,k-1\}$ and $\{k+1,\dots,n\}$, respectively. The notation also covers the additional ties just allowed above. Set

$$u_\varepsilon = \kappa\varepsilon^\delta, \qquad b_\varepsilon = 1+\gamma u_\varepsilon, \qquad d_\varepsilon = 2\varepsilon+\varepsilon^2, \qquad \mathfrak{D}_\varepsilon = d_\varepsilon + u_\varepsilon + \varepsilon u_\varepsilon.$$

Then $Q_1 = \mathcal{D}(1,1,0)$ and $Q_2 = \mathcal{D}\big(1, b_\varepsilon^{-1}, 0\big)$, and

$$H_\varepsilon = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad \begin{aligned} A &= \mathcal{D}(\gamma(1+\varepsilon), \gamma(1+\varepsilon), 1)\,, \\ B &= \mathcal{D}(\gamma,\gamma,0)\,, \\ C &= \mathcal{D}(\gamma,\gamma/b_\varepsilon,0)\,, \\ D &= \mathcal{D}(\gamma(1+\varepsilon), \gamma(1+\varepsilon+u_\varepsilon)/b_\varepsilon, 1)\,. \end{aligned}$$

The exceptional entry of D follows from

$$1 - \frac{1-\gamma-\gamma\varepsilon}{1+\gamma\kappa\varepsilon^\delta} = \frac{\gamma(1+\varepsilon+\kappa\varepsilon^\delta)}{1+\gamma\kappa\varepsilon^\delta}.$$

Block inversion. Write $K = A - BD^{-1}C$. The Schur-complement formula (Cottle, 1974) gives

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} K^{-1} & -K^{-1}BD^{-1} \\ -D^{-1}CK^{-1} & D^{-1} + D^{-1}CK^{-1}BD^{-1} \end{pmatrix}.$$

All four blocks are diagonal, so the products reduce to scalar operations (equivalently, one may permute the coordinates and invert n matrices of size 2×2). First,

$$\begin{aligned} K &= \mathcal{D}(\gamma(1+\varepsilon), \gamma(1+\varepsilon), 1) - \mathcal{D}(\gamma/(1+\varepsilon), \gamma/(1+\varepsilon+u_\varepsilon), 0) \\ &= \mathcal{D}(\gamma d_\varepsilon/(1+\varepsilon), \gamma\mathfrak{D}_\varepsilon/(1+\varepsilon+u_\varepsilon), 1)\,. \end{aligned}$$

Here we used

$$1+\varepsilon - \frac{1}{1+\varepsilon} = \frac{d_\varepsilon}{1+\varepsilon}, \qquad 1+\varepsilon - \frac{1}{1+\varepsilon+u_\varepsilon} = \frac{\mathfrak{D}_\varepsilon}{1+\varepsilon+u_\varepsilon}.$$

In particular, D and K are invertible for $\varepsilon > 0$. Denote the four blocks of H_ε^{-1} by $U_{11}, U_{12}, U_{21}, U_{22}$. Their entries are

$$\begin{aligned} U_{11} = K^{-1} &= \mathcal{D}\left(\frac{1+\varepsilon}{\gamma d_\varepsilon}, \frac{1+\varepsilon+u_\varepsilon}{\gamma\mathfrak{D}_\varepsilon}, 1\right), \\ U_{12} = -K^{-1}BD^{-1} &= -\mathcal{D}\left(\frac{1}{\gamma d_\varepsilon}, \frac{b_\varepsilon}{\gamma\mathfrak{D}_\varepsilon}, 0\right), \\ U_{21} = -D^{-1}CK^{-1} &= -\mathcal{D}\left(\frac{1}{\gamma d_\varepsilon}, \frac{1}{\gamma\mathfrak{D}_\varepsilon}, 0\right). \end{aligned}$$

For the bottom-right block, its entries on $\mathcal{A}$ and at k are, respectively,

$$\begin{aligned} \frac{1}{\gamma(1+\varepsilon)} + \frac{1}{(1+\varepsilon)^2}\frac{1+\varepsilon}{\gamma d_\varepsilon} &= \frac{1+\varepsilon}{\gamma d_\varepsilon}, \\ \frac{b_\varepsilon}{\gamma(1+\varepsilon+u_\varepsilon)} + \frac{b_\varepsilon}{\gamma(1+\varepsilon+u_\varepsilon)}\gamma\frac{1+\varepsilon+u_\varepsilon}{\gamma\mathfrak{D}_\varepsilon}\frac{\gamma}{b_\varepsilon}\frac{b_\varepsilon}{\gamma(1+\varepsilon+u_\varepsilon)} &= \frac{(1+\varepsilon)b_\varepsilon}{\gamma\mathfrak{D}_\varepsilon}. \end{aligned}$$

The latter simplification uses $1+\mathfrak{D}_\varepsilon = (1+\varepsilon)(1+\varepsilon+u_\varepsilon)$. On $\mathcal{I}$ the entry is 1, so

$$U_{22} = D^{-1} + D^{-1}CK^{-1}BD^{-1} = \mathcal{D}\left(\frac{1+\varepsilon}{\gamma d_\varepsilon}, \frac{(1+\varepsilon)b_\varepsilon}{\gamma\mathfrak{D}_\varepsilon}, 1\right).$$

Consequently, the full multiplication giving the Jacobian is

$$\begin{aligned} J_\varepsilon &= -\begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}(-\gamma)\begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix} = \gamma\begin{pmatrix} U_{11}Q_1 + U_{12}Q_2 \\ U_{21}Q_1 + U_{22}Q_2 \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{D}\Big(\dfrac{1+\varepsilon-1}{d_\varepsilon}, \dfrac{1+\varepsilon+u_\varepsilon}{\mathfrak{D}_\varepsilon} - \dfrac{b_\varepsilon}{\mathfrak{D}_\varepsilon}\dfrac{1}{b_\varepsilon}, 0\Big) \\ \mathcal{D}\Big(\dfrac{1+\varepsilon-1}{d_\varepsilon}, \dfrac{(1+\varepsilon)b_\varepsilon}{\mathfrak{D}_\varepsilon b_\varepsilon} - \dfrac{1}{\mathfrak{D}_\varepsilon}, 0\Big) \end{pmatrix} \\ &= \begin{pmatrix} \mathcal{D}\Big(\dfrac{1}{2+\varepsilon}, \dfrac{\varepsilon+u_\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) \\ \mathcal{D}\Big(\dfrac{1}{2+\varepsilon}, \dfrac{\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) \end{pmatrix}. \end{aligned}$$

The entries on $\mathcal{A}$ converge to $1/2$, and those on $\mathcal{I}$ remain zero. The two entries at the designated kink satisfy

$$\begin{aligned} \frac{1}{\mathfrak{D}_\varepsilon}\begin{pmatrix} \varepsilon + u_\varepsilon \\ \varepsilon \end{pmatrix} &= \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} + \frac{1}{2\mathfrak{D}_\varepsilon}\begin{pmatrix} u_\varepsilon - \varepsilon^2 - \varepsilon u_\varepsilon \\ -u_\varepsilon - \varepsilon^2 - \varepsilon u_\varepsilon \end{pmatrix} \\ &\longrightarrow \begin{cases} (1,0)^\top, & \delta < 1 \quad \text{(slow)}, \\ (1+\kappa, 1)^\top/(2+\kappa), & \delta = 1 \quad \text{(critical)}, \\ (1/2, 1/2)^\top, & \delta > 1 \quad \text{(fast)}. \end{cases} \end{aligned}$$

Indeed, $\varepsilon^2 = o(\varepsilon)$ and $\varepsilon u_\varepsilon = o(u_\varepsilon)$, while $u_\varepsilon/\varepsilon = \kappa\varepsilon^{\delta-1}$ distinguishes the three cases. In particular, $\delta = \kappa = 1$ gives the limit $(2/3, 1/3)^\top$. Writing the indicated limit pair as $(j_1, j_2)^\top$, the full limiting Jacobian is

$$J_0 = \begin{pmatrix} \mathcal{D}(1/2, j_1, 0) \\ \mathcal{D}(1/2, j_2, 0) \end{pmatrix}.$$

Within the critical family, letting $\kappa \downarrow 0$ approaches the fast-regime pair and letting $\kappa \to +\infty$ approaches the slow-regime pair. In fact, $\kappa \mapsto (1+\kappa, 1)^\top/(2+\kappa)$ maps $(0, +\infty)$ onto the interior of the segment joining $(1/2, 1/2)^\top$ and $(1, 0)^\top$, and extends continuously to its endpoints.

Verification of the composite residual formula. We now compute the reduced matrices from Section 4.2 of the main paper (Corollary 25). For $Q_\varepsilon = \operatorname{diag}(Q_1, Q_2)$,

$$\begin{aligned} &Q_\varepsilon^\dagger = \operatorname{diag}\big(\mathcal{D}(1,1,0)\,, \mathcal{D}(1, b_\varepsilon, 0)\big), \\ &Q_\varepsilon \longrightarrow Q_0 = \operatorname{diag}\big(\mathcal{D}(1,1,0)\,, \mathcal{D}(1,1,0)\big), \\ &Q_\varepsilon^\dagger \longrightarrow Q_0^\dagger = Q_0 = \Pi_0 = \Pi_\varepsilon. \end{aligned}$$

The projector is independent of ε. Using (SM10),

$$\begin{aligned} \Sigma_\varepsilon &= Q_\varepsilon^\dagger - \Pi_\varepsilon + \gamma\Pi_\varepsilon H^f \Pi_\varepsilon \\ &= \gamma\begin{pmatrix} \mathcal{D}(1,1,0) & \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) & \mathcal{D}(1,1+u_\varepsilon,0) \end{pmatrix} \longrightarrow \Sigma_0 = \gamma\begin{pmatrix} \mathcal{D}(1,1,0) & \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) & \mathcal{D}(1,1,0) \end{pmatrix}. \end{aligned}$$

Since $\Pi_\varepsilon H^f (I - \Pi_\varepsilon) = 0$,

$$\begin{aligned} V_\varepsilon &= Q_\varepsilon^\dagger M_\varepsilon - \gamma\Pi_\varepsilon H^f (I - \Pi_\varepsilon) M_\varepsilon \\ &= Q_\varepsilon^\dagger M_\varepsilon = -\gamma\begin{pmatrix} \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) \end{pmatrix} = V_0. \end{aligned}$$

Next,

$$T_\varepsilon = \Sigma_\varepsilon + \gamma\varepsilon\Pi_\varepsilon = \gamma \begin{pmatrix} \mathcal{D}(1+\varepsilon, 1+\varepsilon, 0) & \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) & \mathcal{D}(1+\varepsilon, 1+\varepsilon+u_\varepsilon, 0) \end{pmatrix} \longrightarrow \Sigma_0.$$

To obtain $W_\varepsilon = \Sigma_\varepsilon^\dagger V_\varepsilon$, the required coordinatewise pseudoinverses are

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}^\dagger = \tfrac{1}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \begin{pmatrix} 1 & 1 \\ 1 & 1+u_\varepsilon \end{pmatrix}^\dagger = \frac{1}{u_\varepsilon}\begin{pmatrix} 1+u_\varepsilon & -1 \\ -1 & 1 \end{pmatrix}.$$

Restoring the original ordering gives the full multiplication

$$\begin{aligned} W_\varepsilon &= \frac{1}{4\gamma}\begin{pmatrix} \mathcal{D}(1, 4(1+u_\varepsilon)/u_\varepsilon, 0) & \mathcal{D}(1, -4/u_\varepsilon, 0) \\ \mathcal{D}(1, -4/u_\varepsilon, 0) & \mathcal{D}(1, 4/u_\varepsilon, 0) \end{pmatrix} (-\gamma) \begin{pmatrix} \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) \end{pmatrix} \\ &= -\tfrac{1}{2}\begin{pmatrix} \mathcal{D}(1,2,0) \\ \mathcal{D}(1,0,0) \end{pmatrix} =: W_0. \end{aligned}$$

Thus W_ε is constant in ε.

For $R_\varepsilon = T_\varepsilon^\dagger \Sigma_\varepsilon$, the required inverses are

$$\begin{pmatrix} 1+\varepsilon & 1 \\ 1 & 1+\varepsilon \end{pmatrix}^\dagger = \frac{1}{d_\varepsilon}\begin{pmatrix} 1+\varepsilon & -1 \\ -1 & 1+\varepsilon \end{pmatrix},$$

$$\begin{pmatrix} 1+\varepsilon & 1 \\ 1 & 1+\varepsilon+u_\varepsilon \end{pmatrix}^\dagger = \frac{1}{\mathfrak{D}_\varepsilon}\begin{pmatrix} 1+\varepsilon+u_\varepsilon & -1 \\ -1 & 1+\varepsilon \end{pmatrix}.$$

Hence, in the original coordinate ordering,

$$T_\varepsilon^\dagger = \frac{1}{\gamma}\begin{pmatrix} \mathcal{D}\Big(\dfrac{1+\varepsilon}{d_\varepsilon}, \dfrac{1+\varepsilon+u_\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) & \mathcal{D}\Big(-\dfrac{1}{d_\varepsilon}, -\dfrac{1}{\mathfrak{D}_\varepsilon}, 0\Big) \\ \mathcal{D}\Big(-\dfrac{1}{d_\varepsilon}, -\dfrac{1}{\mathfrak{D}_\varepsilon}, 0\Big) & \mathcal{D}\Big(\dfrac{1+\varepsilon}{d_\varepsilon}, \dfrac{1+\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) \end{pmatrix}.$$

Multiplication by Σ_ε yields

$$R_\varepsilon = \begin{pmatrix} \mathcal{D}\Big(\dfrac{\varepsilon}{d_\varepsilon}, \dfrac{\varepsilon+u_\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) & \mathcal{D}\Big(\dfrac{\varepsilon}{d_\varepsilon}, \dfrac{\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) \\ \mathcal{D}\Big(\dfrac{\varepsilon}{d_\varepsilon}, \dfrac{\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) & \mathcal{D}\Big(\dfrac{\varepsilon}{d_\varepsilon}, \dfrac{\varepsilon+u_\varepsilon+\varepsilon u_\varepsilon}{\mathfrak{D}_\varepsilon}, 0\Big) \end{pmatrix}.$$

Its limit depends on (δ, κ):

$$R_0 = \begin{pmatrix} \mathcal{D}\big(1/2, r_0^{\mathrm{tl}}, 0\big) & \mathcal{D}\big(1/2, r_0^{\mathrm{d}}, 0\big) \\ \mathcal{D}\big(1/2, r_0^{\mathrm{d}}, 0\big) & \mathcal{D}\big(1/2, r_0^{\mathrm{br}}, 0\big) \end{pmatrix}, \qquad S = R_0 - \Sigma_0^\dagger\Sigma_0 = R_0 - \frac{1}{2\gamma}\Sigma_0.$$

The residual S has nonzero entries only in the two coordinates corresponding to k. The entries of R_0 and S there are

	$(r_0^{\mathrm{tl}}, s_0^{\mathrm{tl}})$	$(r_0^{\mathrm{d}}, s_0^{\mathrm{d}})$	$(r_0^{\mathrm{br}}, s_0^{\mathrm{br}})$
$\delta < 1$	$(1, 1/2)$	$(0, -1/2)$	$(1, 1/2)$
$\delta = 1$	$\left(\dfrac{1+\kappa}{2+\kappa}, \dfrac{\kappa}{4+2\kappa}\right)$	$\left(\dfrac{1}{2+\kappa}, -\dfrac{\kappa}{4+2\kappa}\right)$	$\left(\dfrac{1+\kappa}{2+\kappa}, \dfrac{\kappa}{4+2\kappa}\right)$
$\delta > 1$	$(1/2, 0)$	$(1/2, 0)$	$(1/2, 0)$.

Finally, the projected-inversion decomposition in equation (13) of the main paper gives

$$\begin{aligned}
-H_\varepsilon^{-1}M_\varepsilon &= -(I-\Pi_\varepsilon)M_\varepsilon - T_\varepsilon^\dagger V_\varepsilon \\
&= -(I-\Pi_\varepsilon)M_\varepsilon - T_\varepsilon^\dagger \Sigma_\varepsilon \Sigma_\varepsilon^\dagger V_\varepsilon \\
&= -(I-\Pi_\varepsilon)M_\varepsilon - R_\varepsilon W_\varepsilon \\
&\longrightarrow -(I-\Pi_0)M_0 - R_0 W_0 \\
&= -(I-\Pi_0)M_0 - (\Sigma_0^\dagger\Sigma_0 + S)W_0.
\end{aligned}$$

The pseudoinverse term corresponding to the fast regime is

$$\begin{aligned}
H_0^\dagger &= \frac{1}{4\gamma}\begin{pmatrix} \mathcal{D}(1,1,4\gamma) & \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) & \mathcal{D}(1,1,4\gamma) \end{pmatrix}, \\
-H_0^\dagger M_0 &= \tfrac{1}{2}\begin{pmatrix} \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) \end{pmatrix}.
\end{aligned} \tag{SM11}$$

On the other hand, $(I-\Pi_0)M_0 = 0$ and $\Sigma_0^\dagger\Sigma_0 = \Sigma_0/(2\gamma)$ imply

$$\begin{aligned}
-R_0 W_0 &= \tfrac{1}{2}\begin{pmatrix} \mathcal{D}\big(1, 2r_0^{\mathrm{tl}}, 0\big) \\ \mathcal{D}\big(1, 2r_0^{\mathrm{d}}, 0\big) \end{pmatrix}, \\
-SW_0 &= \begin{pmatrix} \mathcal{D}\big(0, r_0^{\mathrm{tl}} - 1/2, 0\big) \\ \mathcal{D}\big(0, r_0^{\mathrm{d}} - 1/2, 0\big) \end{pmatrix}, \\
-\Sigma_0^\dagger\Sigma_0 W_0 &= \tfrac{1}{2}\begin{pmatrix} \mathcal{D}(1,1,0) \\ \mathcal{D}(1,1,0) \end{pmatrix}.
\end{aligned}$$

The last expression is exactly the pseudoinverse limit in (SM11), completing the verification. For $(\delta,\kappa) = (1,1)$, the critical limiting element differs from the pseudoinverse element by

$$\left\| \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} - \begin{pmatrix} 2/3 \\ 1/3 \end{pmatrix} \right\|_2 = \frac{1}{3\sqrt{2}},$$

whereas the slow limiting element differs by

$$\left\| \begin{pmatrix} 1/2 \\ 1/2 \end{pmatrix} - \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_2 = \frac{1}{\sqrt{2}}.$$

These are the limiting discrepancies illustrated in Figure 5 of the main paper.

SM4 Detailed Convergence Rates

Throughout this section, vector norms are Euclidean and matrix norms are the induced Euclidean operator norms. Fix a parameter θ and suppose that the iterates satisfy the combined Hölderian error estimate of Remark 31 of the main paper,

$$\|x_\varepsilon^k(\theta) - x_{\min}^\star(\theta)\| \le C_1(1-\gamma\varepsilon)^k + C_0\varepsilon^{1/(2p)}, \qquad k \in \mathbb{N}, \tag{SM12}$$

for admissible $0 < \varepsilon \le \varepsilon^*$, where $p \ge 1$, $C_0 > 0$, and $\varepsilon^* > 0$. For each chosen ε, the constants $C_1 > 0$ and $\gamma > 0$ satisfy $0 < \gamma\varepsilon < 1$ and are independent of k; uniformity of C_1 in ε is not required below. The two terms in (SM12) are the approximation error and the regularization error, respectively.

Corollary SM4 (Estimates for ε and k)**.** *Let $\delta > 0$. Under* (SM12)*, choose $0 < \varepsilon \leq \varepsilon^*$ with $0 < \gamma\varepsilon < 1$. If*

$$0 < \varepsilon < \left(\frac{\delta}{C_0}\right)^{2p} \qquad \text{and} \qquad k > \frac{\log((\delta - C_0\varepsilon^{1/(2p)})/C_1)}{\log(1-\gamma\varepsilon)},$$

then $\|x_\varepsilon^k(\theta) - x_{\min}^\star(\theta)\| \leq \delta$.

Proof. The first condition gives $C_0\varepsilon^{1/(2p)} < \delta$, which deals with the regularization error. Since $\log(1-\gamma\varepsilon) < 0$, the second assumption gives

$$k\log(1-\gamma\varepsilon) < \log((\delta - C_0\varepsilon^{1/(2p)})/C_1) \quad \iff \quad C_1(1-\gamma\varepsilon)^k < \delta - C_0\varepsilon^{1/(2p)}.$$

Thus $C_1(1-\gamma\varepsilon)^k + C_0\varepsilon^{1/(2p)} < \delta$, and (SM12) proves the claim. □

This shows that, although the bound need not vanish as $k \to \infty$ for fixed ε, it can be made arbitrarily small by choosing ε sufficiently small and then k sufficiently large.

Consider the Least-Squares problem $\min_{x\in\mathbb{R}^n} \frac{1}{2}\|Ax-\theta\|_2^2$, with $A \in \mathbb{R}^{m\times n}$. Its quadratic error bound has exponent $p = 2$, but the closed form gives $O(\varepsilon)$ Tikhonov bias rather than the generic $O(\varepsilon^{1/4})$ estimate. If $A = 0$, $x_\varepsilon^\star(\theta) = x_{\min}^\star(\theta) = 0$ with zero derivatives. Otherwise, define respectively the operator norm and the minimal nonzero singular value of A by

$$\lambda := \|A\|_{\text{op}}, \qquad \sigma := \min\{\sigma_i(A) : \sigma_i(A) > 0\}.$$

Proposition SM5 (Complete rates for Least-Squares)**.** *Fix $\theta \in \mathbb{R}^m$, $\varepsilon > 0$, and let*

$$D_\varepsilon := (A^\mathsf{T}A + \varepsilon I_n)^{-1}A^\mathsf{T}, \qquad D_0 = A^\dagger.$$

The solution and derivatives satisfy

$$\|x_\varepsilon^\star(\theta) - x_{\min}^\star(\theta)\| \leq \frac{\varepsilon}{\sigma(\sigma^2+\varepsilon)}\|\Pi_{\text{range}(A)}\theta\| \leq \frac{\varepsilon}{\sigma(\sigma^2+\varepsilon)}\|\theta\|,$$
$$\|D_\varepsilon - D_0\| = \frac{\varepsilon}{\sigma(\sigma^2+\varepsilon)} \leq \frac{\varepsilon}{\sigma^3}.$$

Proof. In this setting, the fixed-point mapping reads, for $\varepsilon > 0$ and suitable $\gamma > 0$,

$$\varphi_\varepsilon^\gamma(x,\theta) = (I - \varepsilon\gamma I - \gamma A^\mathsf{T}A)x + \gamma A^\mathsf{T}\theta.$$

Let us first consider the trajectory and show the first inequalities. Recall the closed-form expressions $x_\varepsilon^\star(\theta) = (A^\mathsf{T}A + \varepsilon I_n)^{-1}A^\mathsf{T}\theta$ and $x_{\min}^\star(\theta) = A^\dagger\theta$. Using the singular value decomposition $A = U\Sigma V^\mathsf{T}$, one gets

$$\|x_\varepsilon^\star(\theta) - x_{\min}^\star(\theta)\| = \|V((\Sigma^\mathsf{T}\Sigma + \varepsilon I_n)^{-1}\Sigma^\mathsf{T} - \Sigma^\dagger)U^\mathsf{T}\Pi_{\text{range}(A)}\theta\|.$$

Coordinate-wise, using that $\sigma_i^\dagger = 1/\sigma_i$ if $\sigma_i > 0$ and 0 otherwise,

$$\frac{\sigma_i}{\sigma_i^2+\varepsilon} - \sigma_i^\dagger = \begin{cases} 0, & \text{if } \sigma_i = 0, \\ -\varepsilon/(\varepsilon\sigma_i + \sigma_i^3), & \text{if } \sigma_i > 0. \end{cases}$$

Thus

$$\|x_\varepsilon^\star(\theta) - x_{\min}^\star(\theta)\| \leq \frac{\varepsilon}{\sigma(\sigma^2+\varepsilon)}\|\Pi_{\text{range}(A)}\theta\| \leq \frac{\varepsilon}{\sigma^3}\|\theta\|.$$

In particular, we have convergence in ε^1 instead of the generic estimate $\varepsilon^{1/4}$. For the implicit derivatives, we estimate $\|(A^\mathsf{T}A + \varepsilon I_n)^{-1}A^\mathsf{T} - A^\dagger\|$ using the same singular values. Since $s \mapsto \varepsilon/[s(s^2+\varepsilon)]$ is decreasing for $s > 0$, its maximum over the nonzero singular values is attained at $s = \sigma$. This gives the stated exact operator-norm formula and the same convergence rate. □

References

Heinz Bauschke and Patrick Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. CMS Books in Mathematics. Springer, 01 2017. ISBN 978-3-319-48310-8. doi: 10.1007/978-3-319-48311-5.

Richard W. Cottle. Manifestations of the Schur complement. *Linear Algebra and its Applications*, 8: 189–211, 1974.