Gabriel Peyré
CNRS & DMA, École Normale Supérieure

CHAPTER 11

Theory of Sparse Regularization

September 9, 2026

The theory of sparse recovery asks when a reconstruction is unique, how much noise changes its coefficients, and whether it identifies the correct nonzero locations. A small reconstruction error alone does not guarantee that these locations are recovered. We use convex geometry and dual certificates to establish recovery conditions for the Lasso and its constrained counterpart, then examine their implications for spike deconvolution on increasingly fine grids.

11.1 Existence and Uniqueness

11.1.1 Existence

We study the penalized problem (10.10), with $\lambda > 0$, and its constrained counterpart (10.11):

$$\min_{x \in \mathbb{R}^N} f_\lambda(x) \overset{\text{def.}}{=} \frac{1}{2\lambda}\|y - Ax\|^2 + \|x\|_1 \tag{$\mathcal{P}_\lambda(y)$}$$

and the limiting problem as $\lambda \to 0$,

$$\min_{Ax=y} \|x\|_1 = \min_x f_0(x) \overset{\text{def.}}{=} \iota_{\mathcal{L}_y}(x) + \|x\|_1. \tag{$\mathcal{P}_0(y)$}$$

where $A \in \mathbb{R}^{P \times N}$, and $\mathcal{L}_y \overset{\text{def.}}{=} \{x \in \mathbb{R}^N \;;\; Ax = y\}$.

Recall that we observe noisy measurements

$$y = Ax_0 + w$$

We seek conditions ensuring that x_0 solves $\mathcal{P}_0(Ax_0)$ in the noiseless case, and quantitative bounds on its distance to solutions of $\mathcal{P}_\lambda(Ax_0 + w)$ when λ is chosen according to the noise level.

The objective in ($\mathcal{P}_\lambda(y)$) is continuous and coercive because $\|\cdot\|_1$ is coercive, so a minimizer exists. Uniqueness is not automatic: A may have a nontrivial kernel, and $\|\cdot\|_1$ is not strictly convex. If $y \in \mathrm{Im}(A)$, the constraint set in ($\mathcal{P}_0(y)$) is nonempty. A constrained minimizer also exists but need not be unique.

Figure 11.1 shows how the recovered coefficients change with λ, illustrating the effect of regularization on sparsity and reconstruction accuracy.

11.1.2 Polytope Projection for the Constrained Problem

The next proposition characterizes noiseless ℓ^1 recovery geometrically.

Proposition 11.1. *We denote* $B \overset{\text{def.}}{=} \{x \in \mathbb{R}^N \;;\; \|x\|_1 \leqslant 1\}$. *For* $x_0 \neq 0$, *assuming* $\mathrm{Im}(A) = \mathbb{R}^P$,

$$x_0 \text{ is a solution to } \mathcal{P}_0(Ax_0) \quad \Longleftrightarrow \quad A\frac{x_0}{\|x_0\|_1} \in \partial(AB) \tag{11.1}$$

where "∂" denotes the boundary and $AB = \{Ax \;;\; x \in B\}$.

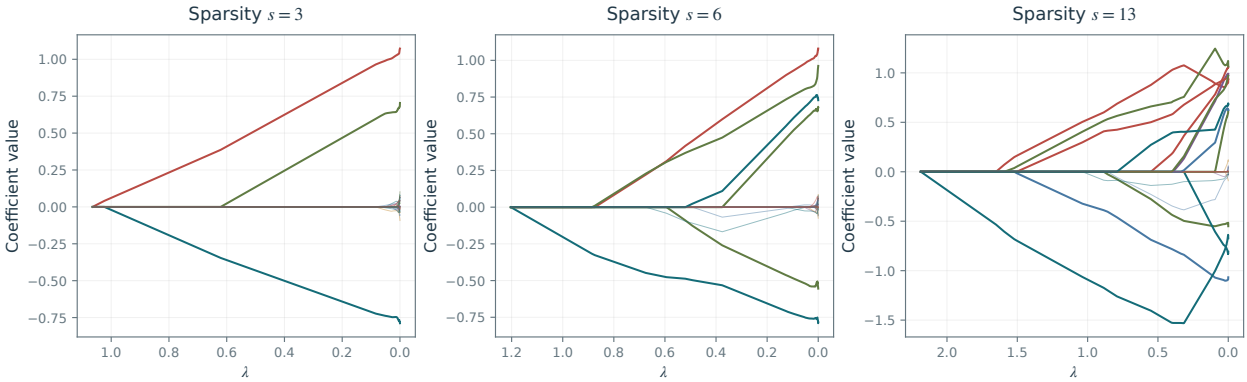

Figure 11.1. Lasso solution paths $\lambda \mapsto x_\lambda$ for sparsities $s = 3, 6, 13$, with the same measurement matrix and noise realization.

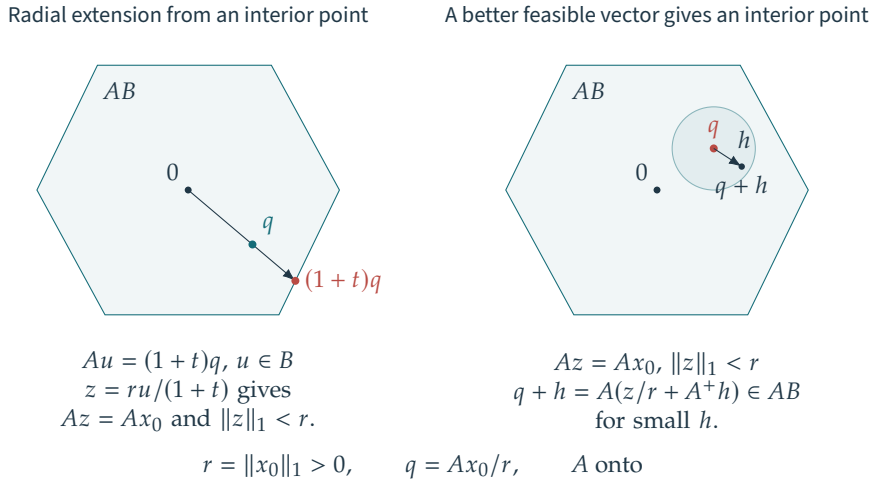

Figure 11.2. Geometry of the polytope characterization of exact recovery.

Proof. Set $r = \|x_0\|_1 > 0$ and $q = Ax_0/r$. If x_0 is not optimal, there is z with $Az = Ax_0$ and $\|z\|_1 < r$. Since A is onto, A^+ is a bounded right inverse. For sufficiently small $h \in \mathbb{R}^P$,

$$\|z/r + A^+h\|_1 \leqslant \|z\|_1/r + \|A^+h\|_1 < 1,$$

and $q + h = A(z/r + A^+h) \in AB$. Thus q is an interior point of AB.

Conversely, if $q \neq 0$ is interior, then $(1+t)q \in AB$ for some $t > 0$. Choose $u \in B$ with $Au = (1+t)q$. The vector $z = ru/(1+t)$ satisfies $Az = Ax_0$ and $\|z\|_1 < r$, so x_0 is not optimal. If $q = 0$, the zero vector is already a better feasible point. These two implications prove the equivalence. □

For a nonzero recovered vector x_0, its measurement Ax_0 lies on the boundary of the scaled polytope $\|x_0\|_1 AB$. When P is much smaller than N, projection can remove faces and prevent ℓ^1 recovery. Chapter 12 studies random projections that preserve enough of the geometry of the ℓ^1 ball in $\mathbb{R}^N$ to allow recovery from $\mathbb{R}^P$.

Identifiability is unchanged by positive scaling: if x_0 is identifiable, so is ρx_0 for every $\rho > 0$. More generally, the relevant face of B, and hence condition (11.1), depends only on $\operatorname{sign}(x_0)$. For this geometric discussion, assume uniqueness and let $\mathcal{A} : y \mapsto x^\star$ map each datum y to the solution of $(\mathcal{P}_0(y))$. By (11.1), A and $\mathcal{A}$ are inverse maps between the cones $C_s = \{x\ ;\ \operatorname{sign}(x) = s\}$ and AC_s for recoverable sign patterns s. Including the lower-dimensional cones and the origin, their images AC_s partition $\mathbb{R}^P$. Suppose that the columns $(a_j)_j$ of A have unit norm. For $P = 3$, intersecting the cones AC_s with the unit sphere in $\mathbb{R}^3$ gives a spherical Delaunay subdivision, which is a triangulation under a general-position assumption; see Figure 11.7. In higher dimensions, simplices replace triangles. The subdivision satisfies the empty-cap property: the spherical cap bounded by a triangle's circumcircle contains no signed column $\pm a_j$ in its interior. Figure 11.3 illustrates these conclusions in $\mathbb{R}^2$ and $\mathbb{R}^3$.

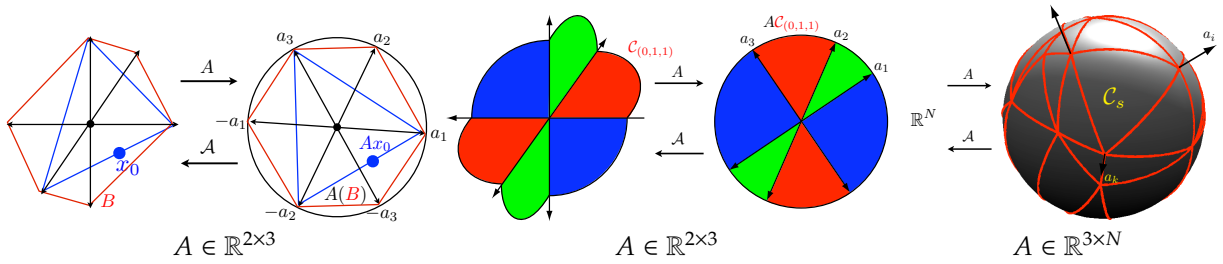

Figure 11.3. The linear map A projects the ℓ^1 ball B; the nonlinear recovery map $\mathcal{A}$ selects solutions of $(\mathcal{P}_0(y))$.

11.1.3 Optimality Conditions

For an index set $I \subset \{1, \dots, N\}$, write $A = (a_i)_{i=1}^N$ for the columns of A and $A_I \stackrel{\text{def.}}{=} (a_i)_{i\in I} \in \mathbb{R}^{P\times|I|}$ for the corresponding submatrix. For a vector $x \in \mathbb{R}^N$, write $x_I \stackrel{\text{def.}}{=} (x_i)_{i\in I} \in \mathbb{R}^{|I|}$ for its restriction.

The following proposition rephrases the first-order optimality conditions in a convenient form.

Proposition 11.2. *x_λ is a solution to $(\mathcal{P}_\lambda(y))$ for $\lambda > 0$ if and only if*

$$\eta_{\lambda,I} = \operatorname{sign}(x_{\lambda,I}) \quad \text{and} \quad \|\eta_{\lambda,I^c}\|_\infty \leqslant 1$$

where we define

$$I \stackrel{\text{def.}}{=} \operatorname{supp}(x_\lambda) \stackrel{\text{def.}}{=} \{i \,;\, x_{\lambda,i} \neq 0\}, \quad \text{and} \quad \eta_\lambda \stackrel{\text{def.}}{=} \frac{1}{\lambda} A^*(y - Ax_\lambda). \tag{11.2}$$

Proof. The objective in $(\mathcal{P}_\lambda(y))$ is the sum of a differentiable convex function and a continuous convex penalty. The subdifferential sum rule gives

$$\partial f_\lambda(x) = \frac{1}{\lambda} A^*(Ax - y) + \partial\|\cdot\|_1(x).$$

Thus x_λ is a solution to $(\mathcal{P}_\lambda(y))$ if and only if $0 \in \partial f_\lambda(x_\lambda)$, which gives the desired result. □

In particular, $\operatorname{supp}(x_\lambda) \subset \operatorname{sat}(\eta_\lambda)$, where $\operatorname{sat}(\eta) = \{i : |\eta_i| = 1\}$ is the saturation set.

For the constrained case $\lambda = 0$, the next proposition introduces dual certificates arising from the Lagrange multipliers associated with $\mathcal{L}_y$.

Proposition 11.3. *$x^\star$ solves $(\mathcal{P}_0(y))$ if and only if $Ax^\star = y$ and*

$$\exists \eta \in \mathcal{D}_0(y, x^\star) \stackrel{\text{def.}}{=} \operatorname{Im}(A^*) \cap \partial\|\cdot\|_1(x^\star). \tag{11.3}$$

Proof. The norm penalty in $(\mathcal{P}_0(y))$ is continuous, so the subdifferential sum rule gives

$$\partial f_0(x) = \partial \iota_{\mathcal{L}_y}(x) + \partial\|\cdot\|_1(x).$$

If $x \in \mathcal{L}_y$, then $\partial \iota_{\mathcal{L}_y}(x)$ is the normal space to the affine set $\mathcal{L}_y$, i.e. $\ker(A)^\perp = \operatorname{Im}(A^*)$. □

Every valid certificate satisfies $\operatorname{supp}(x^\star) \subset \operatorname{sat}(\eta)$, for $\eta \in \mathcal{D}_0(y, x^\star)$.

Writing $I = \operatorname{supp}(x^\star)$, one thus has

$$\mathcal{D}_0(y, x^\star) = \left\{\eta = A^* p \,;\, \eta_I = \operatorname{sign}(x^\star_I),\ \|\eta\|_\infty \leqslant 1\right\}.$$

Although the notation $\mathcal{D}_0(y, x^\star)$ involves a particular solution $x^\star$, the set of certificates is the same for every primal solution. The duality argument below explains this independence.

11.1.4 Uniqueness

The next proposition shows that at least one minimizer uses linearly independent columns of A. It does not assert that every minimizer has this property.

Proposition 11.4. *For $\lambda > 0$, there is a solution $x^\star$ to $(\mathcal{P}_\lambda(y))$ whose support I satisfies $\ker(A_I) = \{0\}$. The same holds for $(\mathcal{P}_0(y))$ when it is feasible.*

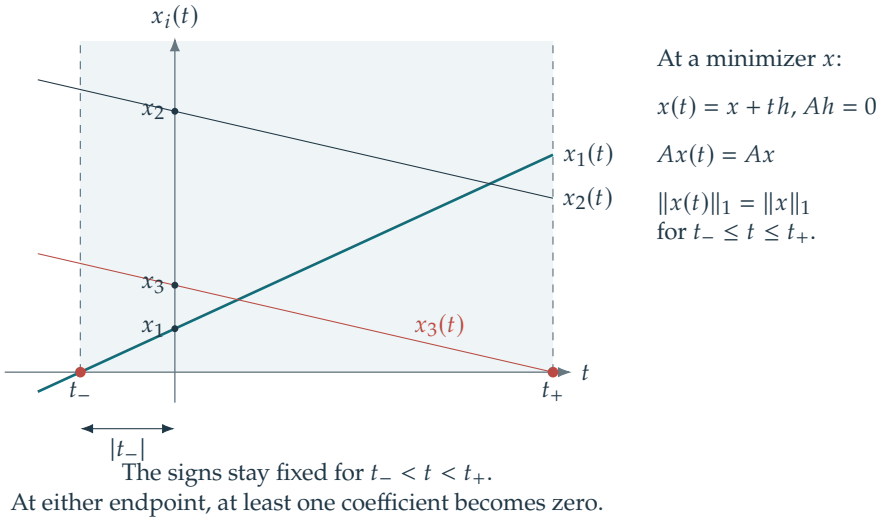

Figure 11.4. Coefficient trajectories along a support-reducing direction.

Proof. Let x be a minimizer and $I = \operatorname{supp}(x)$. If A_I is not injective, choose a nonzero $h_I \in \ker(A_I)$ and extend it by zero outside I. Along the interval containing $t = 0$ on which $x + th$ retains the sign of x, the fidelity is constant and the ℓ^1 norm is affine. Optimality at the interior point $t = 0$ forces its slope to vanish. Move to a finite endpoint of this interval, where at least one coefficient becomes zero. Continuity shows that the endpoint is still a minimizer and has smaller support. Repeating the argument terminates after finitely many steps at a minimizer with injective A_I. □

If the columns of A_I on a minimizer's support are linearly dependent, the proof constructs another minimizer. Thus that solution of ($\mathcal{P}_\lambda(y)$) is nonunique.

For a solution x_λ with $\ker(A_I) = \{0\}$, the optimality condition for ($\mathcal{P}_\lambda(y)$) gives the implicit formula

$$x_{\lambda,I} = A_I^+ y - \lambda (A_I^* A_I)^{-1} \operatorname{sign}(x_{\lambda,I}). \tag{11.4}$$

This expression generalizes soft thresholding (which is recovered when $A = \mathrm{Id}_N$).

Although the coefficient vector x_λ may not be unique, its fitted data Ax_λ, also called the predictor, are unique. Multiplying (11.4) by the restricted measurement matrix gives

$$Ax_\lambda = \operatorname{Proj}_{\operatorname{Im}(A_I)} y - \lambda A_I (A_I^* A_I)^{-1} \operatorname{sign}(x_{\lambda,I}). \tag{11.5}$$

Thus, up to an $O(\lambda)$ bias, this predictor is an orthogonal projection onto a low-dimensional subspace indexed by I.

Lemma 11.5. *For $\lambda > 0$, if x_1, x_2 solve ($\mathcal{P}_\lambda(y)$), then $Ax_1 = Ax_2$. For the feasible constrained problem ($\mathcal{P}_0(y)$), both predictors equal y.*

Proof. For $\lambda = 0$, feasibility gives $Ax_1 = Ax_2 = y$. Otherwise, suppose $Ax_1 \neq Ax_2$ and set $x = (x_1 + x_2)/2$. Convexity gives

$$\|x\|_1 \leqslant \frac{\|x_1\|_1 + \|x_2\|_1}{2} \quad \text{and} \quad \|Ax - y\|^2 < \frac{\|Ax_1 - y\|^2 + \|Ax_2 - y\|^2}{2}$$

where strict convexity of the squared Euclidean norm makes the second inequality strict. Hence

$$\frac{1}{2\lambda}\|Ax - y\|^2 + \|x\|_1 < \frac{1}{2\lambda}\|Ax_1 - y\|^2 + \|x_1\|_1,$$

contradicting the optimality of x_1. □

Proposition 11.6. *For $\lambda > 0$, let x_λ be a solution to ($\mathcal{P}_\lambda(y)$) and denote $\eta_\lambda \overset{\text{def.}}{=} \frac{1}{\lambda} A^*(y - Ax_\lambda)$. We define the "extended support" as*

$$J \overset{\text{def.}}{=} \operatorname{sat}(\eta_\lambda) \overset{\text{def.}}{=} \{i \,;\, |\eta_{\lambda,i}| = 1\}\,.$$

If $\ker(A_J) = \{0\}$ then x_λ is the unique solution of ($\mathcal{P}_\lambda(y)$).

Proof. If $\tilde{x}_\lambda$ is also a minimizer, then by Lemma 11.5, $Ax_\lambda = A\tilde{x}_\lambda$, so that in particular they share the same dual certificate

$$\eta_\lambda = \frac{1}{\lambda}A^*(y - Ax_\lambda) = \frac{1}{\lambda}A^*(y - A\tilde{x}_\lambda).$$

The first-order condition, Proposition 11.2, shows that necessarily $\mathrm{supp}(x_\lambda) \subset J$ and $\mathrm{supp}(\tilde{x}_\lambda) \subset J$. Since $A_J x_{\lambda,J} = A_J \tilde{x}_{\lambda,J}$, and since $\ker(A_J) = \{0\}$, one has $x_{\lambda,J} = \tilde{x}_{\lambda,J}$, and thus $x_\lambda = \tilde{x}_\lambda$ because their supports are included in J. □

Proposition 11.7. *Let $x^\star$ be a solution to ($\mathcal{P}_0(y)$). If there exists $\eta \in \mathcal{D}_0(y, x^\star)$ such that $\ker(A_J) = \{0\}$ where $J \overset{\text{def.}}{=} \mathrm{sat}(\eta)$ then $x^\star$ is the unique solution of ($\mathcal{P}_0(y)$).*

Proof. The proof is the same as for Proposition 11.6, replacing η_λ by η. □

For $\lambda > 0$, a joint density for the entries of A on $\mathbb{R}^{P\times N}$ implies that its columns are in general position almost surely. The Lasso solution is then unique for every y. Randomizing only y, while keeping A fixed, does not in general ensure uniqueness.

11.1.5 Duality

The first-order conditions can also be derived from the convex duality theory in Section 18.3. This interpretation explains why certificates characterize optimal recovery.

Theorem 11.8. *For $\lambda > 0$, strong duality holds between ($\mathcal{P}_\lambda(y)$) and the following dual problem. For $\lambda = 0$, the same statement holds with ($\mathcal{P}_0(y)$) as the primal problem, provided $y \in \mathrm{Im}(A)$:*

$$\sup_{p\in\mathbb{R}^P} \left\{ \langle y,\, p\rangle - \frac{\lambda}{2}\|p\|^2 \;;\; \|A^*p\|_\infty \leqslant 1 \right\}. \tag{11.6}$$

A pair $(x^\star, p^\star)$ is primal–dual optimal if and only if it satisfies the following certificate condition together with the residual relation below:

$$A^*p^\star \in \partial\|\cdot\|_1(x^\star) \quad \Leftrightarrow \quad \left(\|A^*p^\star\|_\infty \leqslant 1, \quad I \subset \mathrm{sat}(A^*p^\star) \quad \textit{and} \quad \mathrm{sign}(x^\star_I) = A^*_I p^\star\right), \tag{11.7}$$

where we denote $I = \mathrm{supp}(x^\star)$, and furthermore, for $\lambda > 0$,

$$p^\star = \frac{y - Ax^\star}{\lambda}.$$

while for $\lambda = 0$, $Ax^\star = y$.

Proof. Introduce $z = Ax$ and define $g_\lambda(z) = \|z - y\|^2/(2\lambda)$ when $\lambda > 0$, and $g_0 = \iota_{\{y\}}$. The Lagrangian is

$$\mathcal{L}(x, z, p) = g_\lambda(z) + \|x\|_1 + \langle z - Ax, p\rangle.$$

For $\lambda > 0$, continuity gives strong duality. For $\lambda = 0$, feasibility and polyhedral convex duality give the same conclusion. Minimizing the Lagrangian over x and z yields

$$\sup_{p\in\mathbb{R}^P} \left[-g_\lambda^*(-p) - (\|\cdot\|_1)^*(A^*p)\right].$$

The conjugates are

$$g_\lambda^*(q) = \langle y, q\rangle + \frac{\lambda}{2}\|q\|^2, \qquad (\|\cdot\|_1)^* = \iota_{\{\|\cdot\|_\infty\leqslant 1\}},$$

including $\lambda = 0$, which gives (11.6). This is also a direct application of the Fenchel–Rockafellar formula (18.17).

The optimality conditions are $z^\star = Ax^\star$, $A^*p^\star \in \partial\|\cdot\|_1(x^\star)$, and $-p^\star \in \partial g_\lambda(z^\star)$. For $\lambda > 0$, the last condition gives $p^\star = (y - Ax^\star)/\lambda$. For $\lambda = 0$, it is equivalent to $Ax^\star = y$. The first condition on $p^\star$ gives the support and sign relations in (11.7). □

For $\lambda > 0$, the dual objective (11.6) is strongly concave. Completing the square shows that its maximizer p_λ is an orthogonal projection:

$$p_\lambda \in \operatorname*{argmin}_{p\in\mathbb{R}^P} \{\|p - y/\lambda\| \;;\; \|A^*p\|_\infty \leqslant 1\}.$$

Thus the supremum is attained at a unique point for $\lambda > 0$. For $\lambda = 0$ and feasible data, a dual optimum also exists, though it need not be unique.

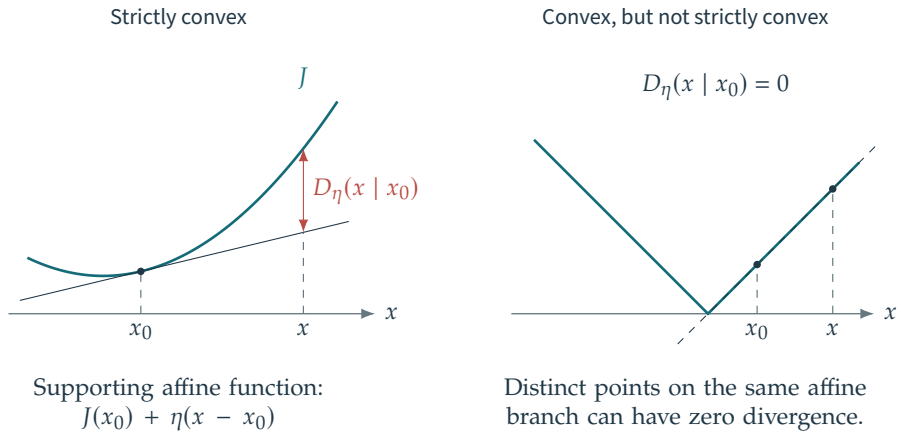

Figure 11.5. Visualization of Bregman divergences.

11.2 Consistency and Sparsistency

11.2.1 Bregman Divergence Rates for General Regularizations

Consider the more general regularized problem

$$\min_{x \in \mathbb{R}^N} \frac{1}{2\lambda}\|Ax - y\|^2 + J(x) \tag{11.8}$$

for a proper convex regularizer J and $\lambda > 0$. We state estimates for any minimizer, without assuming uniqueness.

For x_0 with nonempty subdifferential and a chosen $\eta \in \partial J(x_0)$, define the Bregman divergence

$$D_\eta(x|x_0) \overset{\text{def.}}{=} J(x) - J(x_0) - \langle \eta,\, x - x_0 \rangle.$$

One has $D_\eta(x_0|x_0) = 0$ and, by convexity, $D_\eta(x|x_0) \geqslant 0$; see Figure 11.5.

If J is differentiable, then $\partial J(x_0) = \{\nabla J(x_0)\}$, and the divergence becomes

$$D(x|x_0) \overset{\text{def.}}{=} J(x) - J(x_0) - \langle \nabla J(x_0),\, x - x_0 \rangle.$$

If J is strictly convex, then $D(x|x_0) = 0$ exactly when $x = x_0$. Thus $D(\cdot|\cdot)$ separates points, although it need not be symmetric or satisfy the triangle inequality.

If $J = \|\cdot\|^2$, then $D(x|x_0) = \|x - x_0\|^2$ is the squared Euclidean distance. For the negative-entropy functional $J(x) = \sum_i x_i(\log x_i - 1) + \iota_{\mathbb{R}^N_+}(x)$, one obtains

$$D(x|x_0) = \sum_i x_i \log\left(\frac{x_i}{x_{0,i}}\right) + x_{0,i} - x_i$$

This is the generalized Kullback–Leibler divergence for $x_0 \in \mathbb{R}^N_{++}$ and $x \in \mathbb{R}^N_+$, with the convention $0 \log 0 = 0$. For probability vectors, the linear terms sum to zero.

The following estimate, associated with the work of Burger and Osher, yields a linear noise rate measured by Bregman divergence.

Theorem 11.9. *If there exists*

$$\eta = A^* p \in \operatorname{Im}(A^*) \cap \partial J(x_0), \tag{11.9}$$

then every solution x_λ of (11.8) satisfies

$$D_\eta(x_\lambda|x_0) \leqslant \frac{1}{2}\left(\frac{\|w\|}{\sqrt{\lambda}} + \sqrt{\lambda}\|p\|\right)^2. \tag{11.10}$$

The measurement residual also satisfies

$$\|Ax_\lambda - y\| \leqslant \|w\| + 2\|p\|\lambda. \tag{11.11}$$

Proof. The optimality of x_λ for (11.8) implies

$$\frac{1}{2\lambda}\|Ax_\lambda - y\|^2 + J(x_\lambda) \leqslant \frac{1}{2\lambda}\|Ax_0 - y\|^2 + J(x_0) = \frac{1}{2\lambda}\|w\|^2 + J(x_0).$$

Hence, using $\langle \eta,\, x_\lambda - x_0\rangle = \langle p,\, Ax_\lambda - Ax_0\rangle = \langle p,\, Ax_\lambda - y + w\rangle$, one has

$$\begin{aligned} D_\eta(x_\lambda|x_0) = J(x_\lambda) - J(x_0) - \langle \eta,\, x_\lambda - x_0\rangle &\leqslant \frac{1}{2\lambda}\|w\|^2 - \frac{1}{2\lambda}\|Ax_\lambda - y\|^2 - \langle p,\, Ax_\lambda - y\rangle - \langle p,\, w\rangle \\ &= \frac{1}{2\lambda}\|w\|^2 - \frac{1}{2\lambda}\|Ax_\lambda - y + \lambda p\|^2 + \frac{\lambda}{2}\|p\|^2 - \langle p,\, w\rangle \\ &\leqslant \frac{1}{2\lambda}\|w\|^2 + \frac{\lambda}{2}\|p\|^2 + \|p\|\|w\| = \frac{1}{2}\left(\frac{\|w\|}{\sqrt{\lambda}} + \sqrt{\lambda}\|p\|\right)^2. \end{aligned}$$

Since $D_\eta(x_\lambda|x_0) \geqslant 0$, the completed-square bound gives

$$\|Ax_\lambda - y + \lambda p\|^2 \leqslant \|w - \lambda p\|^2 \leqslant (\|w\| + \lambda\|p\|)^2.$$

The triangle inequality therefore yields

$$\|Ax_\lambda - y\| \leqslant \|w\| + 2\lambda\|p\|.$$

□

When $\|w\| > 0$ and $p \neq 0$, choose $\lambda = \|w\|/\|p\|$ in (11.10). The resulting estimate $D_\eta(x_\lambda|x_0) \leqslant 2\|w\|\|p\|$ is linear in the noise level. For the simple case of a quadratic regularizer $J(x) = \|x\|^2/2$, as used in Section 9.3.2, the source condition (11.9) becomes

$$x_0 \in \mathrm{Im}(A^*)$$

This is (9.12) with $\beta = 1$. Under this condition, (11.10) gives the sublinear ℓ^2 estimate

$$\|x_0 - x_\lambda\| \leqslant 2\sqrt{\|w\|\|p\|}.$$

This agrees with Theorem 9.4 under the convention $x_0 \in \mathrm{Im}((A^*A)^{\beta/2})$.

The source condition (11.9) is sufficient for x_0 to minimize J subject to $Ax = Ax_0$. In finite dimension, it is also necessary under a subdifferential sum-rule qualification, for example when J is finite and continuous on $\mathbb{R}^N$. Without such a qualification, existence of a Lagrange multiplier is an additional assumption.

11.2.2 Linear Rates in Norms for ℓ^1 Regularization

For a regularizer J that is not strictly convex, the Bregman bound (11.10) need not control the error in norm. We now examine how additional certificate conditions restore such control for the ℓ^1 penalty $J = \|\cdot\|_1$.

The next lemma quantifies this control through the gap between $|\eta_i|$ and one.

Lemma 11.10. *For $\eta \in \partial\|\cdot\|_1(x_0)$, let $J \overset{\text{def.}}{=} \mathrm{sat}(\eta)$. Then*

$$D_\eta(x|x_0) \geqslant (1 - \|\eta_{J^c}\|_\infty)\|(x - x_0)_{J^c}\|_1. \tag{11.12}$$

Proof. Note that $x_{0,J^c} = 0$ since $\mathrm{supp}(x_0) \subset \mathrm{sat}(\eta)$ by definition of the subdifferential of the ℓ^1 norm. Since the ℓ^1 norm is separable, each term in the sum defining $D_\eta(x|x_0)$ is nonnegative, hence

$$\begin{aligned} D_\eta(x|x_0) = \sum_i |x_i| - |x_{0,i}| - \eta_i(x_i - x_{0,i}) &\geqslant \sum_{i\in J^c} |x_i| - |x_{0,i}| - \eta_i(x_i - x_{0,i}) \\ = \sum_{i\in J^c} |x_i| - \eta_i x_i &\geqslant \sum_{i\in J^c}(1 - |\eta_i|)|x_i| \geqslant (1 - \|\eta_{J^c}\|_\infty)\sum_{i\in J^c}|x_i| = (1 - \|\eta_{J^c}\|_\infty)\sum_{i\in J^c}|x_i - x_{0,i}|. \end{aligned}$$

□

We use the convention $\|\eta_\emptyset\|_\infty = 0$. The margin $1 - \|\eta_{J^c}\|_\infty > 0$ measures how far η lies from saturation on the complementary coordinates. A larger margin gives stronger control of the coefficient error.

This lemma yields a norm convergence rate under the certificate and injectivity assumptions of Proposition 11.7, with $x^\star = x_0$.

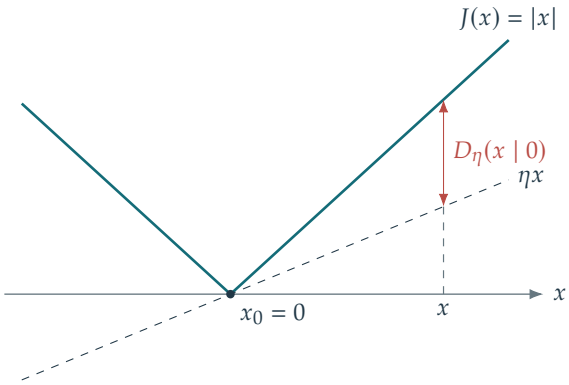

Figure 11.6. Bregman divergence controls the ℓ^1 error on coordinates where η does not saturate.

Theorem 11.11. *If there exists*

$$\eta \in \mathcal{D}_0(Ax_0, x_0) \tag{11.13}$$

and $\ker(A_J) = \{0\}$ *where* $J \overset{\text{def.}}{=} \text{sat}(\eta)$ *then, for fixed* $c > 0$ *and* $\|w\| > 0$, *choosing* $\lambda = c\|w\|$ *gives a constant* C *(depending on* A, *the certificate, and* c, *but not on* w*) such that any solution* x_λ *of* $\mathcal{P}_\lambda(Ax_0 + w)$ *satisfies*

$$\|x_\lambda - x_0\| \leqslant C\|w\|. \tag{11.14}$$

Proof. Write $\eta = A^*p$ and $y = Ax_0 + w$. The optimality of x_λ in ($\mathcal{P}_\lambda(y)$) implies

$$\frac{1}{2\lambda}\|Ax_\lambda - y\|^2 + \|x_\lambda\|_1 \leqslant \frac{1}{2\lambda}\|Ax_0 - y\|^2 + \|x_0\|_1 = \frac{1}{2\lambda}\|w\|^2 + \|x_0\|_1$$

and hence

$$\|Ax_\lambda - y\|^2 \leqslant \|w\|^2 + 2\lambda\|x_0\|_1$$

Using the fact that A_J is injective, one has $A_J^+ A_J = \text{Id}_J$, so that

$$\begin{aligned}
\|(x_\lambda - x_0)_J\|_1 = \|A_J^+ A_J(x_\lambda - x_0)_J\|_1 &\leqslant \|A_J^+\|_{2,1}\|A_J x_{\lambda,J} - y + w\| \leqslant \|A_J^+\|_{2,1}\left(\|A_J x_{\lambda,J} - y\| + \|w\|\right) \\
&\leqslant \|A_J^+\|_{2,1}\left(\|Ax_\lambda - y\| + \|A_{J^c} x_{\lambda,J^c}\| + \|w\|\right) \\
&\leqslant \|A_J^+\|_{2,1}\left(\|Ax_\lambda - y\| + \|A_{J^c}\|_{1,2}\|x_{\lambda,J^c} - x_{0,J^c}\|_1 + \|w\|\right) \\
&\leqslant \|A_J^+\|_{2,1}\left(\|w\| + 2\|p\|\lambda + \|A_{J^c}\|_{1,2}\|x_{\lambda,J^c} - x_{0,J^c}\|_1 + \|w\|\right)
\end{aligned}$$

where we used $x_{0,J^c} = 0$ and (11.11). The mixed operator norm $\|B\|_{p,q}$ maps ℓ^p to ℓ^q, as defined in Remark 11.14. Substituting this bound into the decomposition and using (11.12) and (11.10) gives

$$\begin{aligned}
\|x_\lambda - x_0\|_1 &= \|(x_\lambda - x_0)_J\|_1 + \|(x_\lambda - x_0)_{J^c}\|_1 \\
&\leqslant \|(x_\lambda - x_0)_{J^c}\|_1\left(1 + \|A_J^+\|_{2,1}\|A_{J^c}\|_{1,2}\right) + \|A_J^+\|_{2,1}\left(2\|p\|\lambda + 2\|w\|\right) \\
&\leqslant \frac{D_\eta(x_\lambda|x_0)}{1 - \|\eta_{J^c}\|_\infty}\left(1 + \|A_J^+\|_{2,1}\|A_{J^c}\|_{1,2}\right) + \|A_J^+\|_{2,1}\left(2\|p\|\lambda + 2\|w\|\right) \\
&\leqslant \frac{\frac{1}{2}\left(\frac{\|w\|}{\sqrt{\lambda}} + \sqrt{\lambda}\|p\|\right)^2}{1 - \|\eta_{J^c}\|_\infty}\left(1 + \|A_J^+\|_{2,1}\|A_{J^c}\|_{1,2}\right) + \|A_J^+\|_{2,1}\left(2\|p\|\lambda + 2\|w\|\right).
\end{aligned}$$

Thus setting $\lambda = c\|w\|$, one obtains the constant

$$C \overset{\text{def.}}{=} \frac{\frac{1}{2}\left(\frac{1}{\sqrt{c}} + \sqrt{c}\|p\|\right)^2}{1 - \|\eta_{J^c}\|_\infty}\left(1 + \|A_J^+\|_{2,1}\|A_{J^c}\|_{1,2}\right) + \|A_J^+\|_{2,1}\left(2\|p\|c + 2\right).$$

□

This theorem guarantees uniqueness of the noiseless solution x_0, but does not imply uniqueness of x_λ. The source condition (11.13), combined with the injectivity condition $\ker(A_I) = \{0\}$, gives the ℓ^2 stability estimate (11.14). The estimate (11.14) controls the error linearly relative to the fixed noiseless signal x_0. It does not by itself establish pairwise Lipschitz continuity between arbitrary noisy solutions.

Compare this linear rate with the sublinear rates for quadratic regularization in Theorem 9.4. The source conditions for linear regularization (9.12) and nonlinear regularization (11.13) express different structural assumptions. In the quadratic case with $\beta = 1$, the source condition is $x_0 \in \mathrm{Im}(A^*) = \ker(A)^\perp$ in finite dimension. The signal itself must have no component in $\ker(A)$ because squared-norm regularization cannot recover such a component. The nonlinear condition instead requires a subgradient η to belong to $\mathrm{Im}(A^*)$. It can therefore permit recovery of signal components in $\ker(A)$ when the prior supplies sufficient structure.

11.2.3 Sparsistency for Low Noise

Theorem 11.11 gives an abstract guarantee whose certificate assumptions may be difficult to verify. We now construct a candidate certificate explicitly. When it satisfies a strict off-support bound, it guarantees both a linear rate and support stability.

For any solution x_λ of $(\mathcal{P}_\lambda(y))$, define the dual certificate as in (11.2). Uniqueness of the fitted data makes it independent of the chosen solution:

$$\eta_\lambda \overset{\text{def.}}{=} A^* p_\lambda \quad \text{where} \quad p_\lambda \overset{\text{def.}}{=} \frac{y - Ax_\lambda}{\lambda}.$$

The next proposition shows that p_λ converges to the dual solution of minimum norm for the constrained problem.

Proposition 11.12. *If $y = Ax_0$ where x_0 solves $\mathcal{P}_0(y)$, one has*

$$p_\lambda \to p_0 \overset{\text{def.}}{=} \operatorname*{argmin}_{p \in \mathbb{R}^P} \{\|p\| \;;\; A^* p \in \mathcal{D}_0(y, x_0)\}. \tag{11.15}$$

The vector $\eta_0 \overset{\text{def.}}{=} A^* p_0$ is called the minimum-norm certificate; the minimized norm is that of the data-space vector p_0.

Proof. Write $\mathcal{D} = \{p : \|A^* p\|_\infty \leqslant 1\}$ and let M be the maximum of $\langle y, p\rangle$ on $\mathcal{D}$. Optimality relative to the minimum-norm dual maximizer p_0 gives $0 \leqslant M - \langle y, p_\lambda\rangle \leqslant \frac{\lambda}{2}(\|p_0\|^2 - \|p_\lambda\|^2)$. Thus $\|p_\lambda\| \leqslant \|p_0\|$, every cluster point is a dual maximizer of no greater norm, and uniqueness of p_0 proves convergence as $\lambda \downarrow 0$. □

The certificates in $\mathcal{D}_0(y, x_0)$ for $\lambda = 0$ may be nonunique, but the limit $\lambda \to 0$ selects one of them. This selected certificate governs support stability when both the noise and λ are small.

The inequality constraint $\|\eta_0\|_\infty \leqslant 1$ makes the minimum-norm problem (11.15) difficult to solve explicitly. Dropping this ℓ^∞ constraint while retaining interpolation on the support gives the minimum-norm precertificate:

$$\eta_F \overset{\text{def.}}{=} A^* p_F \quad \text{where} \quad p_F \overset{\text{def.}}{=} \operatorname*{argmin}_{p \in \mathbb{R}^P} \left\{\|p\| \;;\; A_I^* p = \mathrm{sign}(x_{0,I})\right\}. \tag{11.16}$$

The notation "η_F" refers to the "Fuchs" certificate, named after J.-J. Fuchs, who first used it to study ℓ^1 minimization.

The vector p_F may fail dual feasibility because $\|\eta_F\|_\infty \leqslant 1$ need not hold. This is why the construction gives a precertificate. If the interpolation system is consistent, p_F is its minimum-norm solution: $A_I^* p = \mathrm{sign}(x_{0,I})$. The pseudoinverse gives $p_F = A_I^{*,+} \mathrm{sign}(x_{0,I})$; see Proposition 9.2. Under $\ker(A_I) = \{0\}$, this simplifies to

$$p_F = A_I (A_I^* A_I)^{-1} \mathrm{sign}(x_{0,I}).$$

With the Gram matrix $C \overset{\text{def.}}{=} A^* A$, the certificate becomes

$$\eta_F = C_{\cdot,I} C_{I,I}^{-1} \mathrm{sign}(x_{0,I}). \tag{11.17}$$

The next proposition relates η_F to η_0: whenever η_F is feasible, it equals η_0.

Proposition 11.13. *If $\|\eta_F\|_\infty \leqslant 1$, then $p_F = p_0$ and $\eta_F = \eta_0$.*

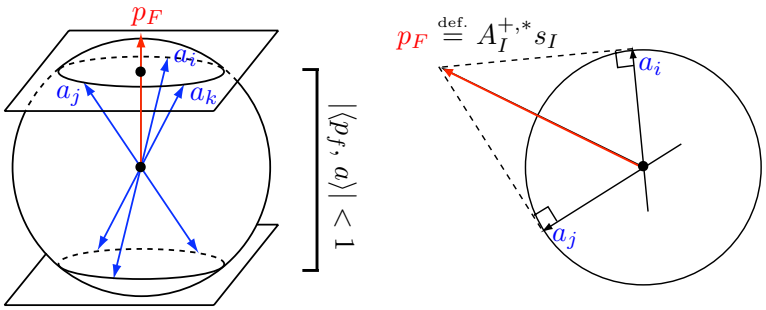

Figure 11.7. For unit-norm columns, the supporting hyperplane defined by a valid certificate determines an empty spherical cap: its interior contains no signed column $\pm a_i$.

The condition $\|\eta_F\|_\infty \leqslant 1$ implies that x_0 is a solution to $(\mathcal{P}_0(y))$. A strict off-support bound on η_F strengthens this recovery condition and gives both a linear error rate and support stability for small noise.

The feasibility condition $\|\eta_F\|_\infty \leqslant 1$ has the empty-cap interpretation shown in Figure 11.7. We thank Charles Dossal for pointing out this connection to spherical Delaunay triangulations.

Remark 11.14 (Operator norm). In the proof, we use the $\ell^p - \ell^q$ matrix operator norm, which is defined as

$$\|B\|_{p,q} \overset{\text{def.}}{=} \max\left\{\|Bu\|_q \;;\; \|u\|_p \leqslant 1\right\}.$$

For $p = q$, we denote $\|B\|_p \overset{\text{def.}}{=} \|B\|_{p,p}$. For $p = q = 2$, $\|B\|_2$ is the maximum singular value, and one has

$$\|B\|_1 = \max_j \sum_i |B_{i,j}| \quad \text{and} \quad \|B\|_\infty = \max_i \sum_j |B_{i,j}|.$$

Theorem 11.15. *Let $I = \operatorname{supp}(x_0) \neq \emptyset$, assume A_I is injective, and suppose $\|\eta_{F,I^c}\|_\infty < 1$. There exist $c_0, c_1 > 0$, depending on A, I, and x_0, such that if*

$$0 < \lambda < c_0, \qquad \|A^* w\|_\infty < c_1 \lambda,$$

then the solution x_λ of $(\mathcal{P}_\lambda(y))$ is unique and has the same support and sign as x_0. It satisfies

$$x_{\lambda,I} = x_{0,I} + A_I^+ w - \lambda (A_I^* A_I)^{-1} \operatorname{sign}(x_{0,I}), \qquad x_{\lambda,I^c} = 0. \tag{11.18}$$

In particular, the error is $O(\|A^* w\|_\infty + \lambda)$. When $\|A^* w\|_\infty > 0$, choosing λ proportional to this quantity with a sufficiently large constant gives an $O(\|w\|)$ error as the noise tends to zero.

Proof. Let $T \overset{\text{def.}}{=} \min_{i \in I} |x_{0,i}|$ be the smallest nonzero signal amplitude, and let $\delta \overset{\text{def.}}{=} \|A^* w\|_\infty$ measure the noise after correlation with the measurement columns. Set $s \overset{\text{def.}}{=} \operatorname{sign}(x_0)$ and use (11.18) to define the candidate

$$\hat{x}_I \overset{\text{def.}}{=} x_{0,I} + A_I^+ w - \lambda (A_I^* A_I)^{-1} s_I, \tag{11.19}$$

and $\hat{x}_{I^c} = 0$. The goal is to show that $\hat{x}$ is indeed the unique solution of $(\mathcal{P}_\lambda(y))$.

Step 1. Sign consistency means $\operatorname{sign}(\hat{x}) = s$. It follows from $\|x_{0,I} - \hat{x}_I\|_\infty < T$, for which a sufficient condition is

$$\|x_{0,I} - \hat{x}_I\|_\infty \leqslant K \|A_I^* w\|_\infty + K\lambda < T \quad \text{where} \quad K \overset{\text{def.}}{=} \|(A_I^* A_I)^{-1}\|_\infty, \tag{11.20}$$

where we used the fact that $A_I^+ = (A_I^* A_I)^{-1} A_I^*$.

Step 2. We verify the strict off-support bound in Proposition 11.6: $\|\hat{\eta}_{I^c}\|_\infty < 1$, where $\lambda \hat{\eta} = A^*(y - A\hat{x})$. Together with sign consistency and injectivity, this proves that $\hat{x}$ is the unique solution of $(\mathcal{P}_\lambda(y))$. Compute

$$\begin{aligned} \lambda \hat{\eta} &= A^*\big(A_I x_{0,I} + w - A_I \hat{x}_I\big) \\ &= A^*(\mathrm{Id} - A_I A_I^+) w + \lambda \eta_F. \end{aligned}$$

The residual formula also gives $\hat{\eta}_I = s_I$. Write $\delta = \|A^* w\|_\infty$, $L = \|A_{I^c}^* A_I\|_\infty$, $R = KL + 1$, and $S = 1 - \|\eta_{F,I^c}\|_\infty > 0$. Then

$$\lambda \|\hat{\eta}_{I^c}\|_\infty \leqslant R\delta + \lambda \|\eta_{F,I^c}\|_\infty < \lambda \quad \text{if} \quad R\delta < S\lambda. \tag{11.21}$$

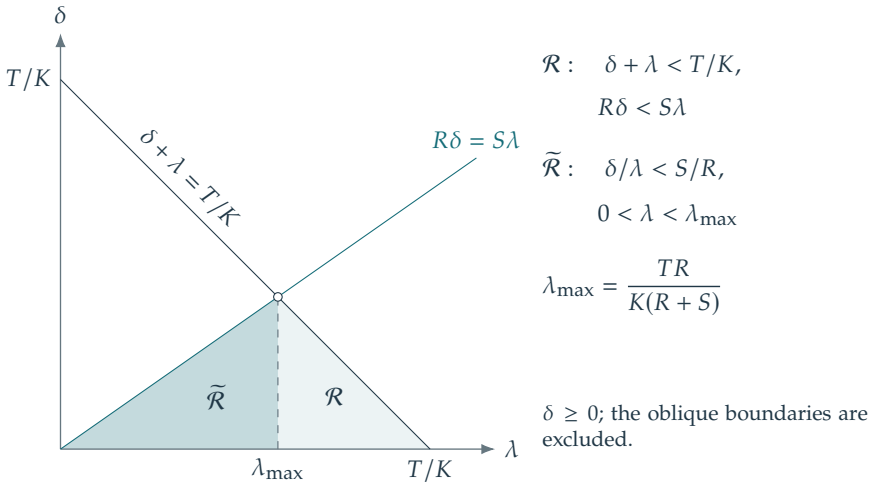

Figure 11.8. Region in the (λ, δ) plane where sign consistency holds.

Together with sign consistency and injectivity of A_I, this strict inequality proves uniqueness.

Conclusion. Combining (11.20) and (11.21), we obtain uniqueness of $\hat{x}$ whenever the parameters (λ, w) satisfy

$$\mathcal{R} = \left\{ (\delta, \lambda) \; ; \; \delta + \lambda < \frac{T}{K} \quad \text{and} \quad R\delta - S\lambda < 0 \right\}$$

The triangular region $\mathcal{R}$ contains the simpler triangle

$$\tilde{\mathcal{R}} = \left\{ (\delta, \lambda) \; ; \; \frac{\delta}{\lambda} < \frac{S}{R} \quad \text{and} \quad \lambda < \lambda_{\max} \right\} \quad \text{where} \quad \lambda_{\max} \overset{\text{def.}}{=} \frac{T(KL+1)}{K(R+S)}. \tag{11.22}$$

□

Theorem 11.15 requires small correlated noise $\|A^* w\|_\infty$ and a small λ, with their ratio controlled, so that regularization does not eliminate the nonzero coefficients of x_0. Theorem 11.11, by contrast, applies at any noise level $\|w\|$, but gives neither support control nor uniqueness of x_λ, and its constants can be less favorable.

The proof provides explicit constants in terms of three parameters K, L, S:

– K accounts for the conditioning of the operator on the support I;

– L is the largest sum of absolute correlations between an off-support atom and all atoms on the support;

– S measures the margin by which η_F avoids saturation outside the support.

The bounds on $\|A^* w\|_\infty / \lambda$ and on λ are given by (11.22). For fixed $\gamma > 0$ and sufficiently small $\delta > 0$, choosing $\lambda = (1+\gamma) R\delta / S$ gives

$$\|x_0 - x_\lambda\|_\infty \leqslant K \left(1 + (1+\gamma) \frac{R}{S} \right) \delta.$$

The strict factor $1 + \gamma$ ensures the required strict inequality. Such a choice is primarily theoretical because its constants involve the unknown support.

For a class of signals x_0, analyzing ℓ^1 support recovery therefore starts with testing the Fuchs precertificate η_F: feasibility requires $\|\eta_F\|_\infty \leqslant 1$, and support stability requires a strict bound outside the support. Figure 11.9 illustrates the behavior for random A: the precertificate η_F becomes nondegenerate when P is sufficiently large. Section 12.2 analyzes this phenomenon.

11.2.4 Sparsistency for Arbitrary Noise

The small-noise theorem fixes the sign of the solution in advance. To obtain support containment for arbitrary noise levels, one can instead control all sign patterns on a prescribed support I. This gives the exact recovery coefficient of Tropp [1]:

$$\mathrm{ERC}(I) = 1 - \max_{j \notin I} \|A_I^+ a_j\|_1.$$

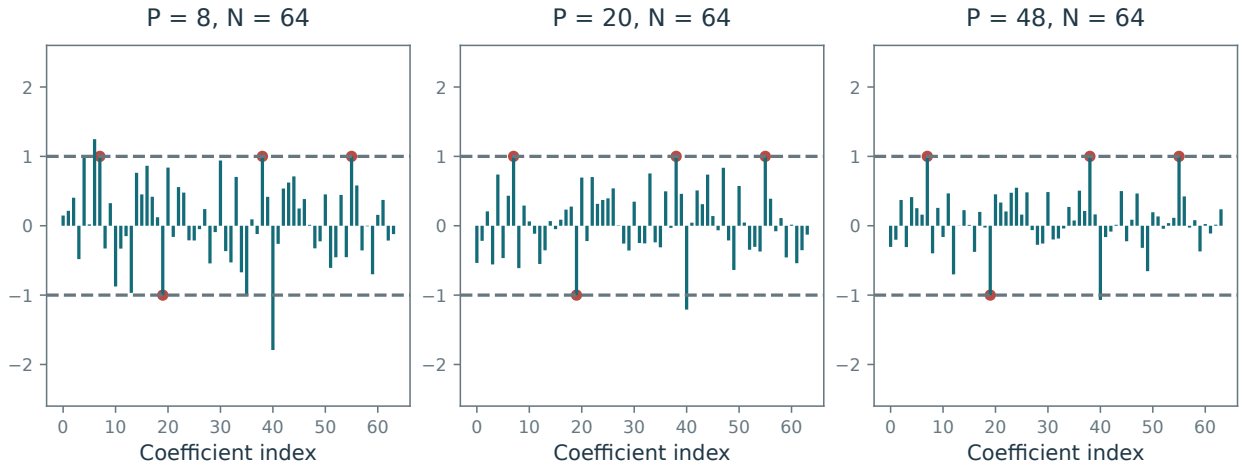

Figure 11.9. Fuchs precertificate η_F for a Gaussian matrix $A \in \mathbb{R}^{P \times N}$ with $N = 64$ and increasing measurement count P.

The maximum over an empty complement is taken to be zero. Assume A_I is injective and $\mathrm{ERC}(I) > 0$. Let $P_I = A_I A_I^+$ and choose $\lambda > 0$ so that

$$\|A_{I^c}^*(\mathrm{Id} - P_I)y\|_\infty < \lambda \, \mathrm{ERC}(I).$$

Then the Lasso has a unique solution supported within I; some coefficients may vanish, so equality of supports is not asserted.

Indeed, minimize the Lasso objective over vectors supported on I. The resulting vector $\hat{x}_I$ is unique, since A_I is injective. Its optimality condition gives $q_I \in \partial\|\cdot\|_1(\hat{x}_I)$ and

$$y - A_I \hat{x}_I = (\mathrm{Id} - P_I)y + \lambda A_I (A_I^* A_I)^{-1} q_I.$$

For $j \notin I$, the corresponding dual coefficient is bounded by

$$\frac{|\langle a_j, (\mathrm{Id} - P_I)y \rangle|}{\lambda} + \|A_I^+ a_j\|_1 < 1.$$

Thus the restricted minimizer satisfies the full optimality conditions with strict inequality outside I, and injectivity proves uniqueness. If $y = Ax_0 + w$ and $\mathrm{supp}(x_0) \subset I$, the orthogonal residual above is $(\mathrm{Id} - P_I)w$. The condition therefore states explicitly how large λ must be relative to the noise.

11.3 Sparse Deconvolution Case Study

For random A, Chapter 12 gives probabilistic conditions under which η_F is a valid certificate.

Super-resolution exposes a limitation of the Fuchs construction. The columns $(a_i)_i$ of A are often samples of a smooth kernel, so nearby columns are highly correlated.

We consider $a_i = \varphi(z_i)$, where $(z_i)_i \subset \mathbb{X}$ is a sampling grid in a domain $\mathbb{X}$ and $\varphi : \mathbb{X} \to \mathcal{H}$ is a smooth map. One has

$$Ax = \sum_i x_i \varphi(z_i).$$

The coefficients of a sparse vector x are the weights of the discrete measure $m_x \overset{\text{def.}}{=} \sum_{i=1}^N x_i \delta_{z_i}$, whose atoms lie on the reconstruction grid.

The matrix A discretizes an operator on Radon measures, $\mathcal{A} : m \in \mathcal{M}(\mathbb{X}) \mapsto y = \mathcal{A}m \in \mathcal{H}$, defined by

$$\mathcal{A}(m) \overset{\text{def.}}{=} \int_{\mathbb{X}} \varphi(x) \mathrm{d}m(x).$$

For a discrete measure, the two representations agree: $\mathcal{A}(m_x) = Ax$.

For example, take $\mathcal{H} = L^2(\mathbb{X})$ on $\mathbb{X} = \mathbb{R}^d$ or $\mathbb{X} = \mathbb{T}^d$ and let $\varphi(z) = \tilde{\varphi}(\cdot - z)$. The resulting operator is convolution:

$$(\mathcal{A}m)(z) = \int \tilde{\varphi}(z - x) \mathrm{d}m(x) = (\tilde{\varphi} \star m)(z).$$

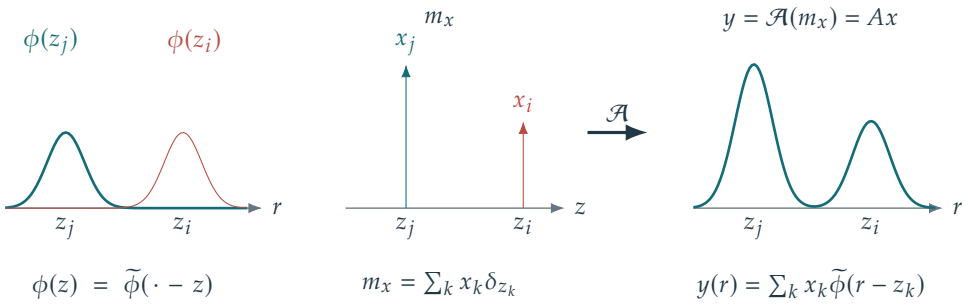

Figure 11.10. Convolution operator.

The observation space $\mathcal{H}$ is infinite-dimensional in this example. Sampling the output instead gives $\varphi(z) = (\tilde{\varphi}(r_j - z))_{j=1}^P$. The observation grid $r \in \mathbb{X}^P$ need not coincide with the reconstruction grid $z \in \mathbb{X}^N$.

A closely related example uses $\varphi(z) = (e^{\mathrm{i}kz})_{k=-f_c}^{f_c}$ on $\mathbb{X} = \mathbb{T}$, so that $\mathcal{A}$ corresponds to computing the $2f_c + 1$ low-frequency coefficients of the Fourier transform of the measure

$$\mathcal{A}(m) = \left(\int_{\mathbb{T}} e^{\mathrm{i}kx} \mathrm{d}m(x) \right)_{k=-f_c}^{f_c}.$$

The operator $\mathcal{A}^*\mathcal{A}$ is a convolution against an ideal low-pass (Dirichlet) kernel. By weighting the Fourier coefficients, one can model low-pass filters on the torus.

On $\mathbb{X} = \mathbb{R}^+$, another example is the Laplace transform:

$$\mathcal{A}(m) = z \mapsto \int_{\mathbb{R}^+} e^{-xz} \mathrm{d}m(x).$$

Define the continuous covariance kernel by

$$\forall\, (z, z') \in \mathbb{X}^2, \quad C(z, z') \overset{\text{def.}}{=} \langle \varphi(z), \varphi(z') \rangle_{\mathcal{H}}.$$

The function C is the kernel of $\mathcal{A}^*\mathcal{A}$. The discrete covariance, defined on the computational grid, is $C = (C(z_i, z_{i'}))_{i,i'} \in \mathbb{R}^{N \times N}$, while its restriction to some support set I is $C_{I,I} = (C(z_i, z_{i'}))_{(i,i') \in I^2} \in \mathbb{R}^{|I| \times |I|}$.

By (11.17), the vector η_F samples the continuous precertificate $\tilde{\eta}_F$ on the grid:

$$\eta_F = (\tilde{\eta}_F(z_i))_{i=1}^N \in \mathbb{R}^N,$$

$$\text{where} \quad \tilde{\eta}_F(x) = \sum_{i \in I} b_i C(x, z_i) \quad \text{where} \quad b_I = C_{I,I}^{-1} \operatorname{sign}(x_{0,I}), \tag{11.23}$$

Thus η_F samples a linear combination of $|I|$ kernel functions $(C(x, z_i))_{i \in I}$.

The question is whether $\|\eta_F\|_{\ell^\infty} \leqslant 1$. Along increasingly dense grids, continuity implies that a uniform bound on the sampled certificate requires $\|\tilde{\eta}_F\|_{L^\infty} \leqslant 1$. A finite grid may miss an overshoot between samples. The construction constrains $\tilde{\eta}_F$ only to interpolate $\operatorname{sign}(x_{0,i})$ at support points z_i; it does not prevent the function from leaving $[-1, 1]$ nearby. Figure 11.11 illustrates this fact.

In one dimension, a certificate bounded in magnitude by one must satisfy $\eta'(z_i) = 0$ at every interior support point $i \in I$, where it already interpolates the sign. Adding these necessary conditions gives the minimum-norm precertificate with vanishing derivatives:

$$\tilde{\eta}_V = \mathcal{A}^* p_V, \qquad p_V = \operatorname*{argmin}_{p \in \mathcal{H}} \left\{ \|p\|_{\mathcal{H}} \,;\, (\mathcal{A}^* p)(z_i) = \operatorname{sign}(x_{0,i}), \quad (\mathcal{A}^* p)'(z_i) = 0 \quad (i \in I) \right\}. \tag{11.24}$$

Here we restrict to a one-dimensional domain and assume that the interpolation constraints are independent. The adjoint is $(\mathcal{A}^* p)(z) = \langle \varphi(z), p \rangle_{\mathcal{H}}$, and $\eta_V = (\tilde{\eta}_V(z_i))_{i=1}^N$. A derivative constraint alone does not ensure $|\tilde{\eta}_V| \leqslant 1$; this bound must be checked separately. As in (11.23), this precertificate is a linear combination of kernel functions, now with $2|I|$ terms:

$$\tilde{\eta}_V(x) = \sum_{i \in I} b_i C(x, z_i) + c_i \partial_2 C(x, z_i),$$

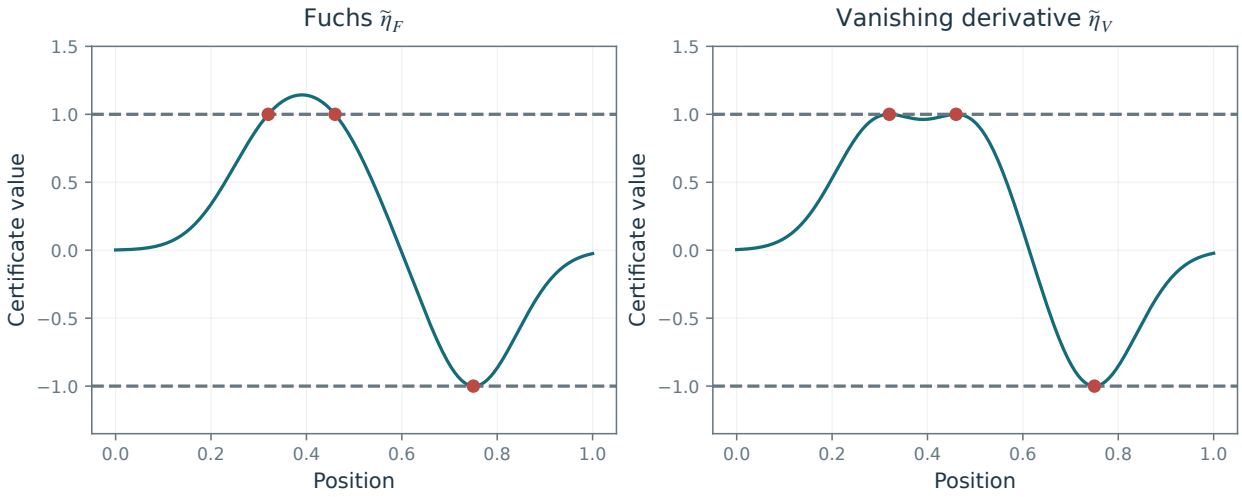

Figure 11.11. Continuous representations of the precertificates η_F and η_V for a convolution operator A.

where $\partial_2 C$ is the derivative of C with respect to the second variable, and (b, c) solve a $2|I| \times 2|I|$ linear system

$$\begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} (C(z_i, z_{i'}))_{i,i' \in I^2} & (\partial_2 C(z_i, z_{i'}))_{i,i' \in I^2} \\ (\partial_1 C(z_i, z_{i'}))_{i,i' \in I^2} & (\partial_1 \partial_2 C(z_i, z_{i'}))_{i,i' \in I^2} \end{pmatrix}^{-1} \begin{pmatrix} \operatorname{sign}(x_{0,I}) \\ \mathbf{0}_I \end{pmatrix}.$$

Evaluating $\tilde{\eta}_V = \mathcal{A}^* p_V$ on the reconstruction grid gives the discrete precertificate η_V. Figure 11.11 shows better behavior of η_V than of η_F. If $\|\eta_V\|_\infty \leqslant 1$ and A is injective on its saturation set, Theorem 11.11 gives a linear rate in the grid-based ℓ^2 norm. As the grid is refined, however, this coefficient ℓ^2 norm does not provide an intrinsic distance between spike measures, which need not have L^2 densities. When η_V differs from η_F, one cannot directly apply Theorem 11.15: exact support stability on increasingly fine grids can fail in super-resolution problems. Methods without a fixed grid provide a more suitable framework for such stability results. In a grid-free formulation, a valid vanishing-derivative precertificate coincides with the minimum-norm certificate under the corresponding interpolation and nondegeneracy assumptions. This leads to stability results for spike locations and amplitudes, rather than a fixed-grid coefficient norm.

References

[1] Joel A. Tropp. Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Transactions on Information Theory*, 52(3):1030–1051, 2006.