Applications

Lijun Zhang
zlj@nju.edu.cn
http://cs.nju.edu.cn/zlj

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Basic Norm Approximation

Norm Approximation Problem

$$\min \quad \|Ax - b\|$$

- $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$ are problem data
- $x \in \mathbf{R}^n$ is the variable
- $\|\cdot\|$ is a norm on $\mathbf{R}^n$
- Approximation solution of $Ax \approx b$, in $\|\cdot\|$

Residual

$$r = Ax - b$$

A Convex Problem

- $b \in \mathcal{R}(A)$, the optimal value is 0
- $b \notin \mathcal{R}(A)$, more interesting ($m > n$)

Basic Norm Approximation

Approximation Interpretation

$$Ax = x_1 a_1 + \cdots + x_n a_n$$

- $a_1, \dots, a_n \in \mathbf{R}^m$ are the columns of A
- Approximate the vector b by a linear combination

- Regression problem
  - ✓ $a_1, \dots, a_n$ are regressors
  - ✓ $x_1 a_1 + \cdots + x_n a_n$ is the regression of b

Basic Norm Approximation

Estimation Interpretation

- Consider a linear measurement model

$$y = Ax + v$$

- $y \in \mathbf{R}^m$ is a vector measurement
- $x \in \mathbf{R}^n$ is a vector of parameters to be estimated
- $v \in \mathbf{R}^m$ is some measurement error that is unknown, but presumed to be small
- Assume smaller values of v are more plausible

$$\hat{x} = \mathrm{argmin}_z \|Az - y\|$$

Basic Norm Approximation

Geometric Interpretation

- Consider the subspace $\mathcal{A} = \mathcal{R}(A) \subseteq \mathbf{R}^m$, and a point $b \in \mathbf{R}^m$
- A projection of the point b onto the subspace $\mathcal{A}$, in the norm $\|\cdot\|$

$$\begin{array}{ll} \min & \|u - b\| \\ \text{s.t.} & u \in \mathcal{A} \end{array}$$

- Parametrize an arbitrary element of $\mathcal{R}(A)$ as $u = Ax$, we see that norm approximation is equivalent to projection

Basic Norm Approximation

Weighted Norm Approximation Problems

$$\min \quad \|W(Ax-b)\|$$

- $W \in \mathbf{R}^{m\times m}$ is called the weighting matrix
  - ✓ The weighting matrix is often diagonal
- A norm approximation problem with norm $\|\cdot\|$, and data $\tilde{A} = WA, \tilde{b} = Wb$
- A norm approximation problem with data A and b, and the W-weighted norm

$$\|z\|_W = \|Wz\|$$

Basic Norm Approximation

Least-Squares Approximation

$$\min \quad \|Ax - b\|_2^2 = r_1^2 + r_2^2 + \cdots + r_m^2$$

- The minimization of a convex quadratic function

$$f(x) = x^\top A^\top Ax - 2b^\top Ax + b^\top b$$

- A point x minimizes f if and only if

$$\nabla f(x) = 2A^\top Ax - 2A^\top b = 0$$

- Normal equations

$$A^\top Ax = A^\top b$$

Basic Norm Approximation

Chebyshev or Minimax Approximation

$$\min \quad \|Ax-b\|_\infty = \max\{|r_1|,\dots,|r_m|\}$$

- Be cast as an LP

$$\begin{array}{ll} \min & t \\ \text{s.t.} & -t\mathbf{1} \preccurlyeq Ax-b \preccurlyeq t\mathbf{1} \end{array}$$

with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}$

Sum of Absolute Residuals Approximation

$$\min \quad \|Ax-b\|_1 = |r_1| + \cdots + |r_m|$$

- Be cast as an LP

$$\begin{array}{ll} \min & \mathbf{1}^{\mathsf{T}} t \\ \text{s.t.} & -t \preccurlyeq Ax-b \preccurlyeq t \end{array}$$

with variables $x \in \mathbf{R}^n$ and $t \in \mathbf{R}^m$

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

l_p-norm Approximation

- l_p-norm approximation, for $1 \leq p \leq \infty$

$$(|r_1|^p + \cdots + |r_m|^p)^{1/p}$$

- The equivalent problem with objective

$$|r_1|^p + \cdots + |r_m|^p$$

  - A separable and symmetric function of the residuals
  - Objective depends only on the amplitude distribution of the residuals

Penalty Function Approximation

The Problem

$$\begin{array}{ll} \min & \phi(r_1) + \cdots + \phi(r_m) \\ \text{s.t.} & r = Ax - b \end{array}$$

- $\phi: \mathbf{R} \to \mathbf{R}$ is called the penalty function
- ϕ is convex
- ϕ is symmetric, nonnegative, and satisfies $\phi(0) = 0$

- A penalty function assesses a cost or penalty for each component of residual

Example

- ℓ_p-norm Approximation

$$\phi(u) = |u|^p$$

  - Quadratic penalty: $\phi(u) = u^2$
  - Absolute value penalty: $\phi(u) = |u|$

- Deadzone-linear Penalty Function

$$\phi(u) = \begin{cases} 0 & |u| \le a \\ |u| - a & |u| > a \end{cases}$$

- The Log Barrier Penalty Function

$$\phi(u) = \begin{cases} -a^2 \log\left(1 - (u/a)^2\right) & |u| < a \\ \infty & |u| \ge a \end{cases}$$

Example

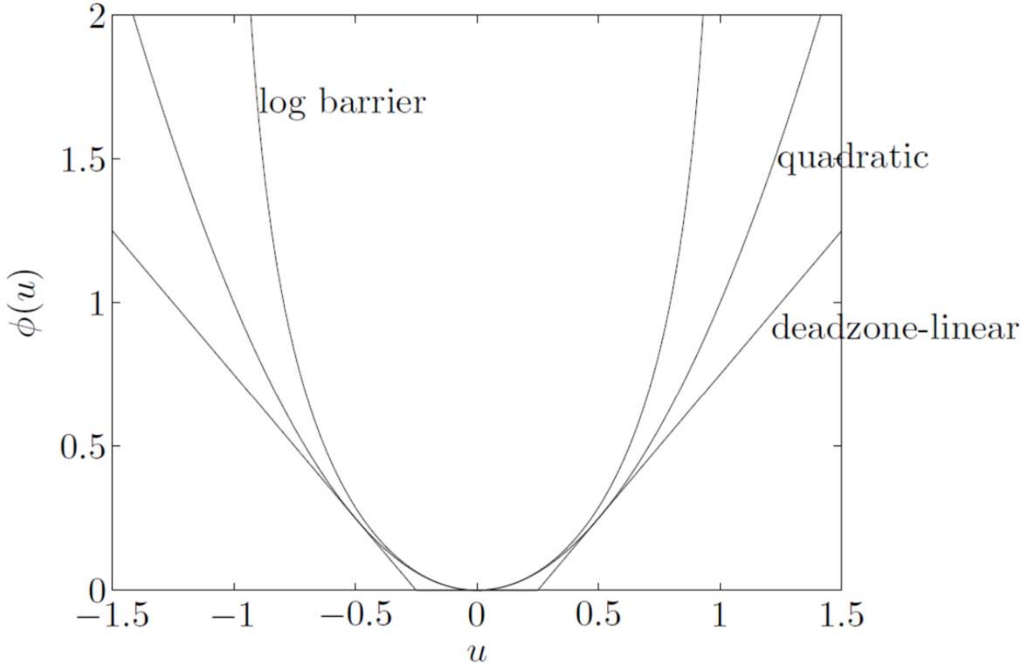


Figure 6.1 Some common penalty functions: the quadratic penalty function $\phi(u) = u^2$, the deadzone-linear penalty function with deadzone width $a = 1/4$, and the log barrier penalty function with limit $a = 1$.

- Log barrier penalty function assesses an infinite penalty for residuals larger than a
- Log barrier function is very close to the quadratic penalty for $|u/a| \leq 0.25$

Discussions

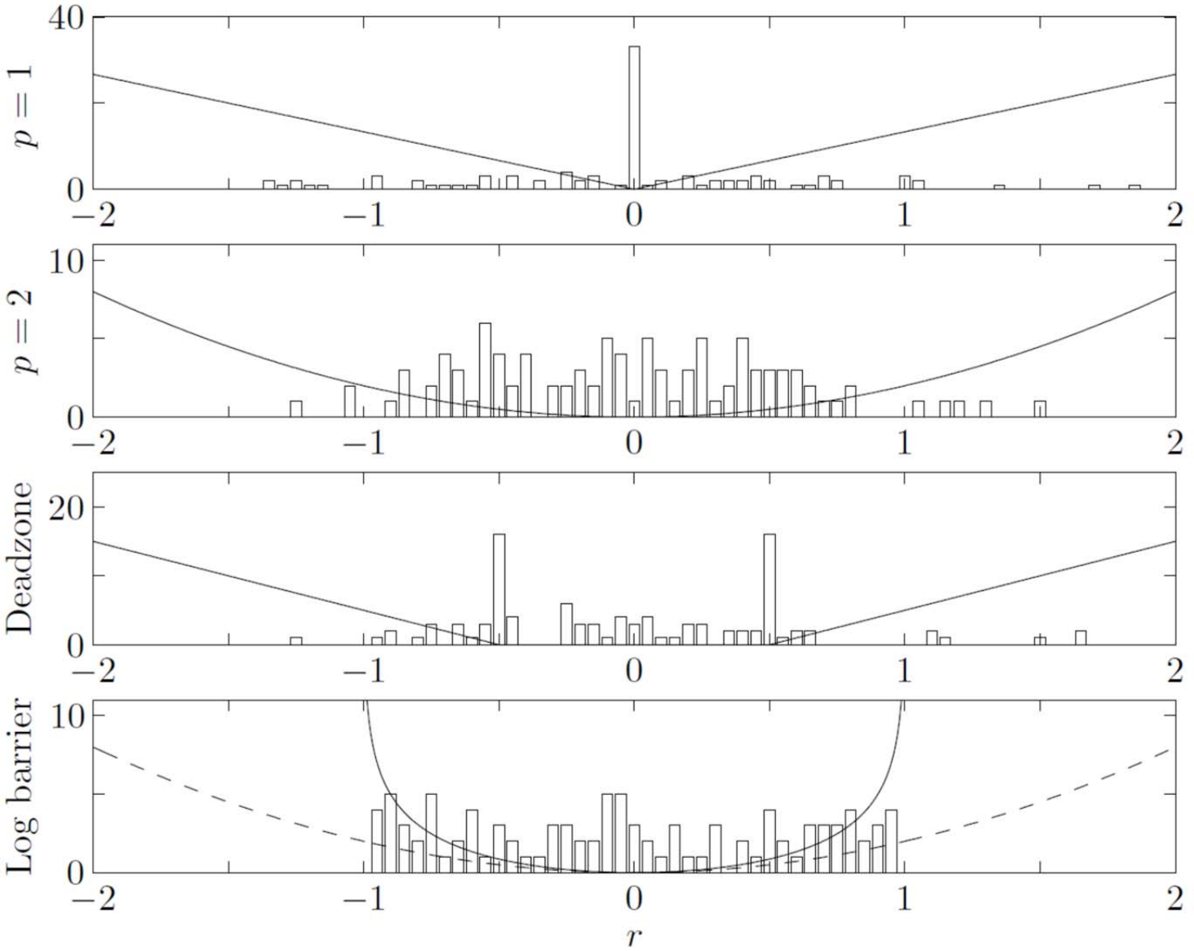
- Roughly speaking, $\phi(u)$ is a measure of our dislike of a residual of value u
- If ϕ is very small for small u, it means we care very little if residuals have these values
- If $\phi(u)$ grows rapidly as u becomes large, it means we have a strong dislike for large residuals
- If ϕ becomes infinite outside some interval, it means that residuals outside the interval are unacceptable

Discussions

- $\phi_1(u) = |u|$、$\phi_2(u) = u^2$
  - For small u we have $\phi_1(u) \gg \phi_2(u)$, so ℓ_1-norm approximation puts relatively larger emphasis on small residuals
  - The optimal residual for the ℓ_1-norm approximation problem will tend to have more zero and very small residuals
  - For large u we have $\phi_2(u) \gg \phi_1(u)$, so ℓ_1-norm approximation puts less weight on large residuals
  - The ℓ_2-norm solution will tend to have relatively fewer large residuals

Example

- $A \in \mathbf{R}^{100\times 30}$, $b \in \mathbf{R}^{100}$

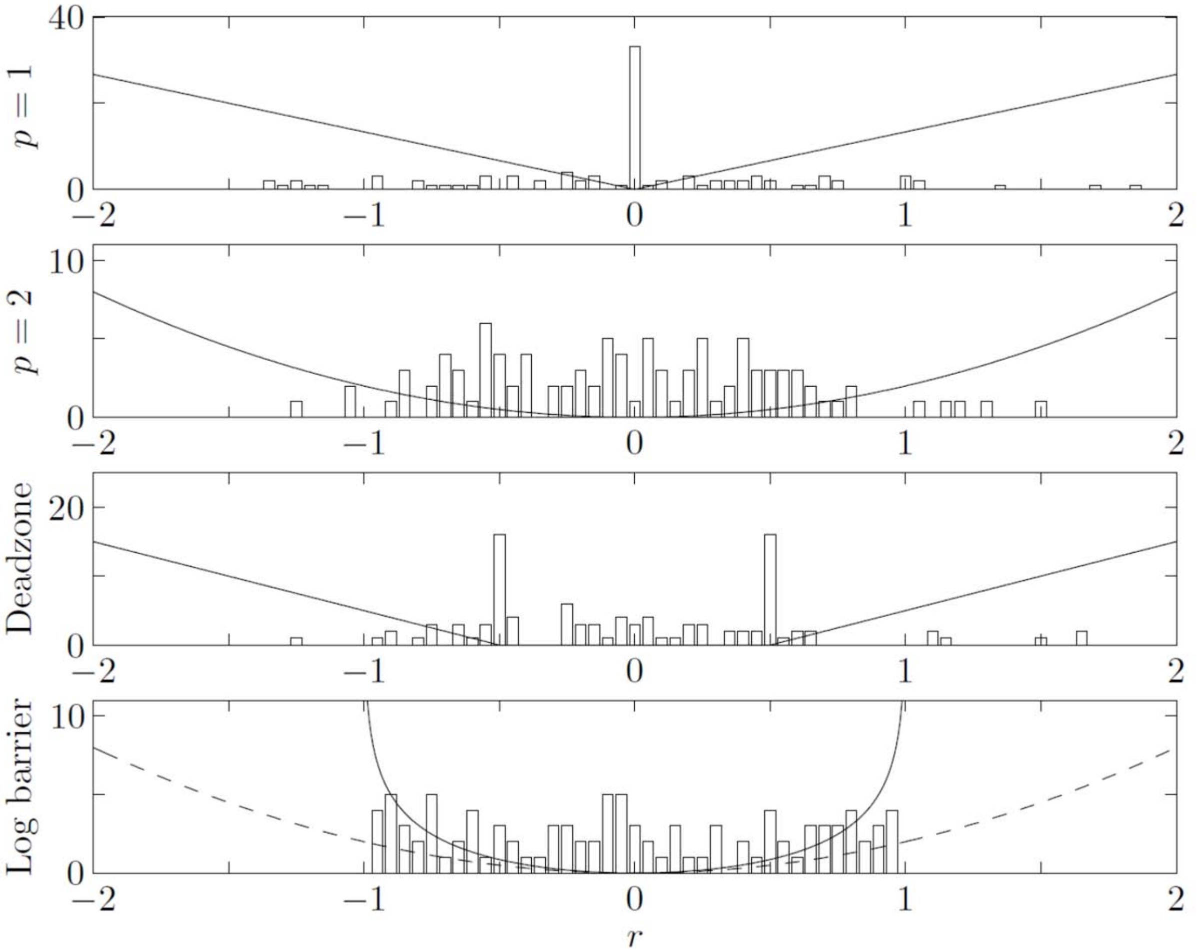


Observations of Penalty Functions

- The ℓ_1-norm penalty puts the most weight on small residuals and the least weight on large residuals.
- The ℓ_2-norm penalty puts very small weight on small residuals, but strong weight on large residuals.
- The deadzone-linear penalty function puts no weight on residuals smaller than 0.5, and relatively little weight on large residuals.
- The log barrier penalty puts weight very much like the ℓ_2-norm penalty for small residuals, but puts very strong weight on residuals larger than around 0.8, and infinite weight on residuals larger than 1.

Example

- $A \in \mathbf{R}^{100\times 30}$, $b \in \mathbf{R}^{100}$



Observations of Amplitude Distributions

- For the ℓ_1-optimal solution, many residuals are either zero or very small. The ℓ_1-optimal solution also has relatively more large residuals.
- The ℓ_2-norm approximation has many modest residuals, and relatively few larger ones.
- For the deadzone-linear penalty, we see that many residuals have the value ± 0.5, right at the edge of the 'free' zone, for which no penalty is assessed.
- For the log barrier penalty, we see that no residuals have a magnitude larger than 1, but otherwise the residual distribution is similar to the residual distribution for ℓ_2-norm approximation.

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Approximation with Constraints

- Add Constraints to

$$\min \quad \|Ax - b\|$$

  - Rule out certain unacceptable approximations of the vector b
  - Ensure that the approximator Ax satisfies certain properties
  - Prior knowledge of the vector x to be estimated
  - Prior knowledge of the estimation error v
  - Determine the projection of a point b on a set more complicated than a subspace

Approximation with Constraints

- Nonnegativity Constraints on Variables

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & x \succcurlyeq 0 \end{aligned}$$

  - Estimate a vector x of parameters known to be nonnegative
  - Determine the projection of a vector b onto the cone generated by the columns of A
  - Approximate b using a nonnegative linear combination of the columns of A

Approximation with Constraints

Variable Bounds

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & l \preccurlyeq x \preccurlyeq u \end{aligned}$$

- Prior knowledge of intervals in which each variable lies

- Determine the projection of a vector b onto the image of a box under the linear mapping induced by A

Approximation with Constraints

□ Probability Distribution

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & x \succcurlyeq 0, \mathbf{1}^{\mathsf{T}} x = 1 \end{aligned}$$

- Estimation of proportions or relative frequencies
- Approximate b by a convex combination of the columns of A

□ Norm Ball Constraint

$$\begin{aligned} \min \quad & \|Ax - b\| \\ \text{s.t.} \quad & \|x - x_0\| \leq d \end{aligned}$$

- x_0 is prior guess of what the parameter x is, and d is the maximum plausible deviation

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Least-norm Problems

Basic least-norm Problem

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

- $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$
- $x \in \mathbf{R}^n$, $\|\cdot\|$ is a norm on $\mathbf{R}^n$
- The solution is called a least-norm solution of $Ax = b$
- A convex optimization problem
- Interesting when $m < n$
  - ✓ When the equation is underdetermined

Least-norm Problems

- Reformulation as Norm Approximation Problem
  - Let x_0 be any solution of $Ax = b$
  - Let $Z \in \mathbf{R}^{n\times k}$ be a matrix whose columns are a basis for the nullspace of A.

$$\{x|Ax = b\} = \{x_0 + Zu|u \in \mathbf{R}^k\}$$

  - The least-norm problem can be expressed as

$$\min \quad \|x_0 + Zu\|$$

Least-norm Problems

Estimation Interpretation

- We have $m < n$ perfect linear measurement, given by $Ax = b$
- Our measurements do not completely determine x
- Suppose our prior information, is that x is more likely to be small than large
- Choose the parameter vector x which is smallest among all parameter vectors that are consistent with the measurements

Least-norm Problems

Geometric Interpretation

- The feasible set $\{x|Ax = b\}$ is affine
- The objective is the distance between x and the point 0

- Find the point in the affine set with minimum distance to 0
- Determine the projection of the point 0 on the affine set $\{x|Ax = b\}$

Least-norm Problems

□ Least-squares Solution of Linear Equations

$$\begin{aligned} \min \quad & \|x\|_2^2 \\ \mathrm{s.t.} \quad & Ax = b \end{aligned}$$

- The optimality conditions

$$2x^* + A^\top v^* = 0 \quad Ax^* = b$$

  - ✓ v is the dual variable

- The Solution

$$x^* = -\frac{1}{2}A^\top v^* \Rightarrow -\frac{1}{2}AA^\top v^* = b$$

$$\Rightarrow \quad v^* = -2(AA^\top)^{-1}b, x^* = A^\top(AA^\top)^{-1}b$$

Least-norm Problems

Least-penalty Problems

$$\begin{array}{ll} \min & \phi(x_1) + \cdots + \phi(x_n) \\ \text{s.t.} & Ax = b \end{array}$$

- $\phi: \mathbf{R} \to \mathbf{R}$ is convex, nonnegative and satisfies $\phi(0) = 0$
- The penalty function value $\phi(u)$ quantifies our dislike of a component of x having value u
- Find x that has least total penalty, subject to the constraint $Ax = b$

Least-norm Problems

□ Sparse Solutions via Least ℓ_1-norm

$$\begin{array}{ll} \min & \|x\|_1 \\ \text{s.t.} & Ax = b \end{array}$$

- Tend to produce a solution x with a large number of components equal to 0
- Tend to produce sparse solutions of $Ax = b$, often with m nonzero components

Least-norm Problems

- Sparse Solutions via Least ℓ_1-norm

$$\begin{array}{ll} \min & \|x\|_1 \\ \text{s.t.} & Ax = b \end{array}$$

- Find solutions of $Ax = b$ that have only m nonzero components
  - $\tilde{A}$ is a submatrix of A
  - $\tilde{x}$ and subvector of x
  - Solve $\tilde{A}\tilde{x} = b$
    - ✓ If there is a solution, we are done
  - Complexity: $n!/(m!\,(n-m)!)$

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Bi-criterion Formulation

- A (convex) Vector Optimization Problem with Two Objectives

$$\min(\text{w.r.t. } \mathbf{R}_+^2) \quad (\|Ax-b\|, \|x\|)$$

  - Find a vector x that is small
  - Make the residual $Ax-b$ small
  - Optimal trade-off between the two objectives
    - ✓ The minimum value of $\|x\|$ is 0 and the residual norm is $\|b\|$
    - ✓ Let C denote the set of minimizers of $\|Ax-b\|$, and then any minimum norm point in C is Pareto optimal

Regularization

- Weighted Sum of the Objectives

$$\min \quad \|Ax - b\| + \gamma\|x\|$$

  - $\gamma > 0$ is a problem parameter
  - A common scalarization method used to solve the bi-criterion problem
  - As γ varies over $(0, \infty)$, the solution traces out the optimal trade-off curve

- Weighted Sum of Squared Norms

$$\min \quad \|Ax - b\|^2 + \gamma\|x\|^2$$

Regularization

Tikhonov Regularization

$$\min \quad \|Ax - b\|_2^2 + \delta\|x\|_2^2 = x^\top(A^\top A + \delta I)x - 2b^\top Ax + b^\top b$$

- Analytical solution

$$x = (A^\top A + \delta I)^{-1}A^\top b$$

- Since $A^\top A + \delta I \succ 0$ for any $\delta \succ 0$, the Tikhonov regularized least-squares solution requires no rank assumptions on the matrix A

Regularization

- ℓ_1-norm Regularization

$$\min \quad \|Ax - b\|_2 + \gamma \|x\|_1$$

  - Find a sparse solution
  - The residual is measured with the Euclidean norm and the regularization is done with an ℓ_1-norm
  - By varying γ we can sweep out the optimal trade-off curve between $\|Ax - b\|_2$ and $\|x\|_1$
    - ✓ As an approximation of the optimal trade-off curve between $\|Ax - b\|_2$ and the cardinality $\mathrm{card}(x)$ of the vector x

Example

Regressor Selection Problem

$$\begin{array}{ll} \min & \|Ax - b\|_2 \\ \text{s.t.} & \operatorname{card}(x) \le k \end{array}$$

- One straightforward approach is to check every possible sparsity pattern in x with k nonzero entries
- For a fixed sparsity pattern, we can find the optimal x by solving a least-squares problem
- Complexity: $n!/(k!\,(n-k)!)$

Example

Regressor Selection Problem

$$\begin{aligned} \min \quad & \|Ax - b\|_2 \\ \text{s.t.} \quad & \mathrm{card}(x) \le k \end{aligned}$$

- A good heuristic approach is to solve the following problem for different γ

$$\min \quad \|Ax - b\|_2 + \gamma \|x\|_1$$

- Find the smallest value of γ that results in a solution with $\mathrm{card}(x) = k$
- We then fix this sparsity pattern and find the value of x that minimizes $\|Ax - b\|_2$

Example

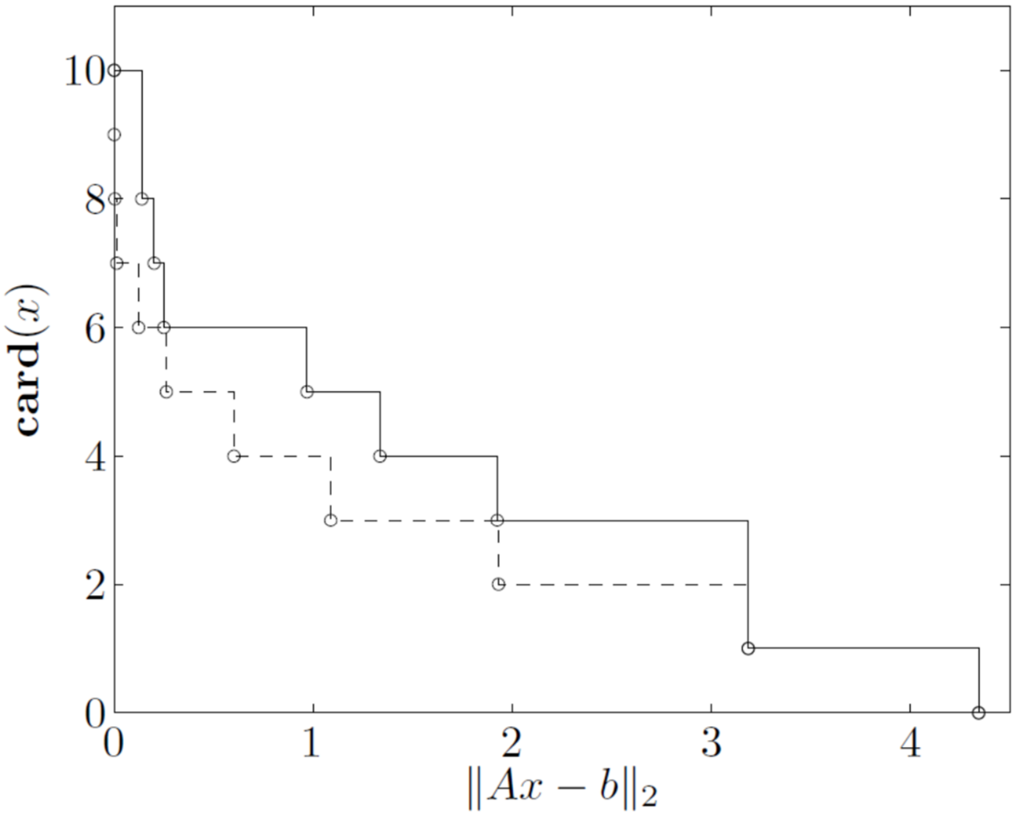


Figure 6.7 Sparse regressor selection with a matrix $A \in \mathbf{R}^{10\times 20}$. The circles on the dashed line are the Pareto optimal values for the trade-off between the residual $\|Ax - b\|_2$ and the number of nonzero elements $\mathbf{card}(x)$. The points indicated by circles on the solid line are obtained via the ℓ_1-norm regularized heuristic.

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Projection on a Set

- The distance of a point $x_0 \in \mathbf{R}^n$ to a closed set $C \subseteq \mathbf{R}^n$, in the norm $\|\cdot\|$

$$\mathrm{dist}(x_0, C) = \inf\{\|x_0 - x\| \,|\, x \in C\}$$

  - The infimum is always achieved
- Projection of x_0 on C
  - Any point $z \in C$ which is closest to x_0

$$\|z - x_0\| = \mathrm{dist}(x_0, C)$$

  - Can be more than one projection of x_0 on C
  - If C is closed and convex, and the norm is strictly convex, there is exactly one

Projection on a Set

- The distance of a point $x_0 \in \mathbf{R}^n$ to a closed set $C \subseteq \mathbf{R}^n$, in the norm $\|\cdot\|$

$$\operatorname{dist}(x_0, C) = \inf\{\|x_0 - x\| \mid x \in C\}$$

  - The infimum is always achieved

- $P_C : \mathbf{R}^n \longrightarrow \mathbf{R}^n$ to denote the projection of x_0 on C

$$P_C(x_0) \in C, \|x_0 - P_C(x_0)\| = \operatorname{dist}(x_0, C)$$

$$P_C(x_0) = \operatorname{argmin}\{\|x - x_0\| \mid x \in C\}$$

  - We refer to P_C as projection on C

Example

- Projection on the Unit Square in $\mathbf{R}^2$
  - Consider the boundary of the unit square in $\mathbf{R}^2$, i.e., $C = \{x \in \mathbf{R}^2 \mid \|x\|_\infty = 1\}$, take $x_0 = 0$
  - In the ℓ_1-norm, the four points $(1,0)$, $(0,-1)$, $(-1,0)$, and $(0,1)$ are closest to $x_0 = 0$, with distance 1, so we have $\text{dist}(x_0, C) = 1$ in the ℓ_1-norm
  - In the ℓ_∞-norm, all points in C lie at a distance 1 from x_0, and $\text{dist}(x_0, C) = 1$

Example

- Projection onto Rank-k Matrices
  - The set of $m \times n$ matrices with rank less than or equal to k

$$C = \{X \in \mathbf{R}^{m \times n} | \operatorname{rank} X \leq k\}$$

    with $k \leq \min\{m, n\}$
  - The Projection of $X_0 \in \mathbf{R}^{m \times n}$ on C in $\|\cdot\|_2$
    - ✓ SVD of X_0

$$X_0 = \sum_{i=1}^{r} \sigma_i u_i v_i^\top$$

$$P_C(x_0) = \sum_{i=1}^{\min\{k,r\}} \sigma_i u_i v_i^\top$$

Outline

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set

Projection on a Convex Set

- C is Convex
  - Represent C by a set of linear equalities and convex inequalities

$$Ax = b, \qquad f_i(x) \le 0, i = 1, \dots, m$$

- Projection of x_0 on C

$$\begin{aligned} \min \quad & \|x - x_0\| \\ \text{s.t.} \quad & f_i(x) \le 0, i = 1, \dots, m \\ & Ax = b \end{aligned}$$

  - A convex optimization problem
  - Feasible if and only if C is nonempty

Example

Euclidean Projection on a Polyhedron

- Projection of x_0 on $C = \{x | Ax \preccurlyeq b\}$

$$\begin{aligned} \min \quad & \|x - x_0\|_2^2 \\ \text{s.t.} \quad & Ax \preccurlyeq b \end{aligned}$$

- Projection of x_0 on $C = \{x | a^\top x = b\}$

$$P_C(x_0) = x_0 + \frac{(b - a^\top x_0)a}{\|a\|_2^2}$$

- Projection of x_0 on $C = \{x | a^\top x \leq b\}$

$$P_C(x_0) = \begin{cases} x_0 + \dfrac{(b - a^\top x_0)a}{\|a\|_2^2}, & a^\top x_0 > b \\ x_0, & a^\top x_0 \leq b \end{cases}$$

Example

Euclidean Projection on a Polyhedron

- Projection of x_0 on $C = \{x | l \preccurlyeq x \preccurlyeq u\}$

$$P_C(x_0)_k = \begin{cases} l_k, & x_{0k} \le l_k \\ x_{0k}, & l_k \le x_{0k} \le u_k \\ u_k, & u_k \le x_{0k} \end{cases}$$

Property of Euclidean Projection

- C is Convex

$$\|P_C(x) - P_C(x)\|_2 \le \|x - y\|_2$$

for all x, y

Example

Euclidean Projection on a Proper Cone

- Projection of x_0 on a proper cone K

$$\begin{aligned} \min \quad & \|x - x_0\|_2^2 \\ \text{s.t.} \quad & x \succcurlyeq_K 0 \end{aligned}$$

- KKT Conditions

$$x \succcurlyeq_K 0, \qquad x - x_0 = z, \qquad z \succcurlyeq_{K^*} 0, \qquad z^\top x = 0$$

- Introduce $x_+ = x$ and $x_- = z$

$$x_0 = x_+ - x_-, \qquad x_+ \succcurlyeq_K 0, \qquad x_- \succcurlyeq_{K^*} 0, \qquad x_+^\top x_- = 0$$

- Decompose x_0 into two orthogonal elements
  - ✓ One nonnegative with respect to K
  - ✓ The other nonnegative with respect to K^*

Example

- $K = \mathbf{R}_+^n$

$$P_K(x_0)_k = \max\{x_{0k}, 0\}$$

  - Replace each negative component with 0
- $K = \mathbf{S}_+^n$ and $\|\cdot\|_F$

$$P_K(X_0) = \sum_{i=1}^{n} \max\{0, \lambda_i\}\, v_i v_i^\top$$

  - The eigendecomposition of X_0 is $X_0 = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$
  - Drop terms associated with negative eigenvalues

Summary

- Norm Approximation
  - Basic Norm Approximation
  - Penalty Function Approximation
  - Approximation with Constraints
- Least-norm Problems
- Regularized Approximation
- Projection
  - Projection on a Set
  - Projection on a Convex Set