Two Dual Problems to ℓ_1-Regularized Least Squares

Mark Schmidt
Department of Computer Science
University of British Columbia

December 4, 2008

Abstract

We derive two problems that are dual the problem of minimizing the squared error in a linear regression model, with a penalty on the ℓ_1-norm of the coefficients.

Dual Problem A: p variables, bound constraints

The primal problem is

$$\min_{\mathbf{x}} \frac{1}{2}||A\mathbf{x} - \mathbf{b}||_2^2 + \lambda||\mathbf{x}||_1.$$

We introduce a dummy variables $\mathbf{y}$ into the ℓ_1-norm, along with a set of trivial equality constraints:

$$\min_{\mathbf{x},\mathbf{y}} \frac{1}{2}||A\mathbf{x} - \mathbf{b}||_2^2 + \lambda||\mathbf{y}||_1 \quad s.t. \quad \mathbf{y} = \mathbf{x}.$$

Using $\mathbf{z}$ to denote the Lagrange multipliers, we can write the Lagrangian of this problem as

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{z}) \triangleq \frac{1}{2}||A\mathbf{x} - \mathbf{b}||_2^2 + \lambda||\mathbf{y}||_1 + \mathbf{z}^T(\mathbf{x} - \mathbf{y}).$$

Distributing $\mathbf{z}$ across the subtraction and grouping terms involving $\mathbf{x}$ and $\mathbf{y}$, the resulting dual function is

$$\max_{\mathbf{z}} \inf_{\mathbf{x},\mathbf{y}} \frac{1}{2}||A\mathbf{x} - \mathbf{b}||_2^2 + \mathbf{z}^T\mathbf{x} + \lambda||\mathbf{y}||_1 - \mathbf{z}^T\mathbf{y}. \tag{1}$$

We first simplify this expression by computing the infimum over $\mathbf{x}$. The derivatives of the Lagrangian with respect to $\mathbf{x}$ are

$$\nabla_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = A^T(A\mathbf{x} - \mathbf{b}) + \mathbf{z},$$
$$\nabla^2_{\mathbf{x}}\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{z}) = A^TA.$$

Equating the first derivative with $\mathbf{0}$, we obtain that $\mathbf{x}$ solves the system

$$A^TA\mathbf{x} = A^T\mathbf{b} - \mathbf{z}.$$

This stationary point is a global minima because the second derivative is positive semi-definite for all $\mathbf{x}$. Assuming that A has independent columns, the optimal $\mathbf{x}$ in terms of $\mathbf{z}$ is the unique solution

$$\mathbf{x} = (A^TA)^{-1}(A^T\mathbf{b} - \mathbf{z}). \tag{2}$$

We now consider computing the infimum of over $\mathbf{y}$ for the terms involving $\mathbf{y}$ in (1). Using the definition of the conjugate function to the ℓ_1-norm [see Boyd and Vandenberghe, 2004], we get

$$\inf_{\mathbf{y}} \lambda||\mathbf{y}||_1 - \mathbf{z}^T\mathbf{y} = -\sup_{\mathbf{y}} \mathbf{z}^T\mathbf{y} - \lambda||\mathbf{y}||_1 = \begin{cases} \mathbf{0} & \text{if} \quad ||\mathbf{z}||_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases} \tag{3}$$

We now plug in (2) and (3) into (1) to get

$$\max_{\mathbf{z}:||\mathbf{z}||_\infty \leq \lambda} \frac{1}{2}||A(A^TA)^{-1}(A^T\mathbf{b} - \mathbf{z}) - \mathbf{b}||_2^2 + \mathbf{z}^T(A^TA)^{-1}(A^T\mathbf{b} - \mathbf{z}).$$

Now its time to simplify this monster. We will use $q(\mathbf{z})$ as the term inside the max, and use B to denote $(A^TA)^{-1}$. Expanding out terms, we get

$$q(\mathbf{z}) = \frac{1}{2}(A^T\mathbf{b} - \mathbf{z})^TB^TA^TAB(A^T\mathbf{b} - \mathbf{z}) - (A^T\mathbf{b} - \mathbf{z})^TB^TA^T\mathbf{b} + \mathbf{b}^T\mathbf{b} + \mathbf{z}^TBA^T\mathbf{b} - \mathbf{z}^TB\mathbf{z}.$$

After removing the term that does not depend on $\mathbf{z}$, we use $B = B^T$ and use $(A^TA)B = I$ to get

$$q(\mathbf{z}) = \frac{1}{2}(A^T\mathbf{b} - \mathbf{z})^TB(A^T\mathbf{b} - \mathbf{z}) - (A^T\mathbf{b} - \mathbf{z})^TBA^T\mathbf{b} + \mathbf{z}^TBA^T\mathbf{b} - \mathbf{z}^TB\mathbf{z}.$$

Now expand some more to get

$$q(\mathbf{z}) = \frac{1}{2}\mathbf{b}^TABA^T\mathbf{b} - \mathbf{b}^TAB\mathbf{z} + \frac{1}{2}\mathbf{z}^TB\mathbf{z} - \mathbf{b}^TABA^T\mathbf{b} + \mathbf{z}^TBA^T\mathbf{b} + \mathbf{z}^TBA^T\mathbf{b} - \mathbf{z}^TB\mathbf{z}.$$

Use that $\mathbf{z}^TBA^T\mathbf{b} = \mathbf{b}^TAB\mathbf{z}$, remove terms not involing $\mathbf{z}$, and add/subtract terms to finally get

$$q(\mathbf{z}) = \mathbf{z}^TBA^T\mathbf{b} - \frac{1}{2}\mathbf{z}^TB\mathbf{z}.$$

Note that $BA^T\mathbf{b} = \mathbf{x}_{LS}$ (the least squares estimate), so the dual problem simplifies to

$$\max_{\mathbf{z}:||\mathbf{z}||_\infty \leq \lambda} \mathbf{z}^T\mathbf{x}_{LS} - \frac{1}{2}\mathbf{z}^TB\mathbf{z}.$$

We can write this as a quadratic program with bound constraints,

$$\min_{\mathbf{z}} \quad \frac{1}{2}\mathbf{z}^TB\mathbf{z} - \mathbf{z}^T\mathbf{x}_{LS}, \quad s.t. \quad \forall_i - \lambda \leq z_i \leq \lambda.$$

The optimal primal solution is given by (2). An interpretation of the dual is that it is finding a sub-gradient of the scaled ℓ_1-norm term that turns the primal problem into a simple quadratic minimization problem.

In Matlab (for small problems only):

```
z = quadprog(inv(A'*A),-A\b,[],[],[],[],-lambda*ones(p,1),lambda*ones(p,1));
x = (A'*A)\(A'*b - z);
```

Dual Problem B: n variables, 2p linear constraints

The primal problem is again

$$\min_{\mathbf{x}} \frac{1}{2}||A\mathbf{x} - \mathbf{b}||_2^2 + \lambda||\mathbf{x}||_1.$$

We introduce a dummy variables $\mathbf{r}$ into the ℓ_2-norm, along with a set of equality constraints:

$$\min_{\mathbf{x},\mathbf{r}} \frac{1}{2}||\mathbf{r}||_2^2 + \lambda||\mathbf{x}||_1 \quad s.t. \quad \mathbf{r} = A\mathbf{x} - \mathbf{b}.$$

Using $\mathbf{z}$ to denote the Lagrange multipliers, we can write the Lagrangian of this problem as

$$\mathcal{L}(\mathbf{x}, \mathbf{r}, \mathbf{z}) \triangleq \frac{1}{2}||\mathbf{r}||_2^2 + \lambda||\mathbf{x}||_1 + \mathbf{z}^T(A\mathbf{x} - \mathbf{b} - \mathbf{r}).$$

Distributing $\mathbf{z}$ across the subtraction and grouping terms involving $\mathbf{x}$ and $\mathbf{r}$, the resulting dual function is

$$\max_{\mathbf{z}} \inf_{\mathbf{x},\mathbf{r}} \mathbf{z}^T A\mathbf{x} + \lambda||\mathbf{x}||_1 + \frac{1}{2}||\mathbf{r}||_2^2 - \mathbf{z}^T\mathbf{r} - \mathbf{z}^T\mathbf{b}. \tag{4}$$

We fist simplify this expression by computing the infimum over $\mathbf{x}$ for terms involving $\mathbf{x}$. Using the definition of the conjugate function to the ℓ_1-norm [see Boyd and Vandenberghe, 2004], we get

$$\inf_{\mathbf{x}} \mathbf{z}^T A\mathbf{x} + \lambda||\mathbf{x}||_1 = -\sup_{\mathbf{x}} -\mathbf{z}^T A\mathbf{x} - \lambda||x||_1 = \begin{cases} \mathbf{0} & \text{if} \quad ||A^T\mathbf{z}||_\infty \leq \lambda \\ -\infty & \text{otherwise} \end{cases} \tag{5}$$

We next simplify (4) by computing the infimum over $\mathbf{r}$ for terms involving $\mathbf{r}$. Using the definition of the conjugate function to the ℓ_2-norm squared [see Boyd and Vandenberghe, 2004], we get

$$\inf_{\mathbf{r}} \frac{1}{2}||\mathbf{r}||_2^2 - \mathbf{z}^T\mathbf{r} = -\sup_{\mathbf{r}} \mathbf{z}^T\mathbf{r} - \frac{1}{2}||\mathbf{r}||_2^2 = -\frac{1}{2}\mathbf{z}^T\mathbf{z}. \tag{6}$$

We now plug (5) and (6) into (4) to get

$$\max_{\mathbf{z}} -\frac{1}{2}\mathbf{z}^T\mathbf{z} - \mathbf{z}^T\mathbf{b}, \quad s.t. \quad ||A^T\mathbf{z}||_\infty \leq \lambda.$$

This can be written as a quadratic program with a diagonal second-order term

$$\min_{\mathbf{z}} \frac{1}{2}\mathbf{z}^T\mathbf{z} + \mathbf{z}^T\mathbf{b}, \quad s.t. \quad \lambda \leq A^T\mathbf{z} \leq \lambda.$$

In Matlab (for small problems only):

```
[n,p] = size(A);
z = quadprog(eye(n),y,[X';-X'],lambda*ones(2*p,1));
x = (A'*A)\(A'*b - A'*z);
```

References

S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.