

Relations between β and δ for QP and LP in Compressed Sensing Computations

Jun Zhang, Jun Wang, and Guangwu Xu
Department of Electrical Engineering and Computer Science
University of Wisconsin-Milwaukee
Milwaukee, WI 53211

Abstract

In many compressed sensing applications, linear programming (LP) has been used to reconstruct a sparse signal. When observation is noisy, the LP formulation is extended to allow an inequality constraint and the solution is dependent on a parameter δ , related to the observation noise level. Recently, some researchers also considered quadratic programming (QP) for compressed sensing signal reconstruction and the solution in this case is dependent on a Lagrange multiplier β . In this work, we investigated the relation between δ and β and derived an upper and a lower bound on β in terms of δ . For a given δ , these bounds can be used to approximate β . Since δ is a physically related quantity and easy to determine for an application while there is no easy way in general to determine β , our results can be used to set β when the QP is used for compressed sensing. Our results and experimental verification also provide some insight into the solutions generated by compressed sensing.

1. Introduction

In many compressed sensing applications, signal reconstruction is carried out by solving a linear programming (LP) problem [1]. When observation is noiseless, the LP problem is:

$$(\text{LP}) : \quad \min \|u\|_1 \quad \text{subject to } Au = b \quad (1)$$

where u is the signal to be reconstructed (n -dim vector, sufficiently sparse), A is an $m \times n$ sampling matrix, with $m \ll n$, b is the observation (an m -dim vector), and the minimization is with respect to u . When the observation is noisy (with bounded noise), the LP formulation is extended to

$$(\text{LPn}) : \quad \min \|u\|_1 \quad \text{subject to } \|Au - b\|_2^2 \leq \delta^2 \quad (2)$$

where δ^2 is a bound on noise power, or *noise level*. Although strictly speaking this is no longer a linear programming problem (it is still a convex optimization problem), because its relation to eqn (??) and for the sake of simplicity, we will still refer to it as a part of the LP problem (LPn).

In some work, a quadratic programming (QP) problem, with

$$(\text{QP}) : \quad \min \left\{ \frac{1}{2} \|Au - b\|_2^2 + \beta \|u\|_1 \right\} \quad (3)$$

has been considered for compressed sensing. For example, Fuchs [2, 3] and Troop [4] used the QP formulation to study the theoretical properties of the solutions of various compressed sensing problems. Similarly, Chen et al. (e.g., see [5]) used a QP-type formulation in several papers for compressed sensing based 3D CT (computer tomography) [in their case, the l_1 norm is replaced by total variation (TV)]. Finally, the QP problem is closely related to (in fact equivalent to) the Lasso procedure [6] which is widely used in statistics, pattern recognition, and data mining.

Given the LP and QP formulations, a natural question is: when are they the same, i.e., producing the same results? When the observation is noiseless, Fuchs [2] showed that the QP becomes the same as the LP when $\beta \rightarrow 0^+$ (i.e., from the right). When the observation is noisy, Fuchs [3] pointed out that for a given noise bound δ^2 in the LPn, there exists a $\beta > 0$ for the QP such that the resulting QP is the same as the LPn, i.e., they produce the same solutions; in fact, this can be established through the theory of duality [7]. However, Fuchs also mentioned that it is difficult to find an explicit (e.g., an analytic) relation between this β and the δ^2 in LPn. This means that when a QP algorithm is used in practice for compressed sensing, such as in the previously mentioned 3D CT and Lasso applications, it may not actually be performing compressed sensing (as defined by the LPn) since, in these applications, it is unclear how to find the “right” β . Indeed, in practice β is often selected experimentally. However, as illustrated in Fig. 1 (figures are all at the end of the paper), for a given compressed sensing problem with a given noise level (δ^2), the solution from the QP is dependent on parameter β and most of the time, it is not the same as that from the LPn (this is illustrated through the QP’s reconstruction error for various β values).

In this paper, we attempt to find an analytic relation between β and δ^2 (or equivalently, between β and δ). Such a relation is useful in three respects. First, it may allow us to gain more insight into the relation between the LP and the QP problems, thereby more insight into the nature of the solution of the compressed sensing problem. Second, if in practice we want to use the QP or Lasso for some reason (e.g., familiarity, easier implementation, or faster speed) as an algorithm for compressed sensing, we can obtain or estimate the appropriate β from δ , which, as the noise level, is a physical parameter and usually readily available. Finally, many signal/image processing and computer vision problems are solved by a Bayesian formulation where an energy function related to the posterior probability distribution is minimized. Usually, the energy function is a sum of two terms: the first is related to the observation model and is similar to, or the same as, the first term in QP [see eqn (3)] and the second is related to prior constraints and is similar to the second term in the QP. The two terms are “balanced” by a parameter β just like that in the QP formulation. In many such Bayesian applications, selecting the value of β is a problem and there is no analytic/theoretical guidance. Our work sheds light on this problem and could potentially be used to find a solution.

The rest of the paper is organized as follows. In Section 2, we derive some analytic relations between β and δ and in Section 3, we verify and illustrate some of these relations experimentally. Finally, in Section 4, we provide conclusions.

2. Analytic Results

In this section, we derive two relations between β and δ , one in inequalities and the other in an equality.

2.1. Inequality Relations

Suppose for a noise level or bound δ^2 , the LPn problem of eqn (2) has a sparse solution. Then, as described in Section 1, there is a $\beta > 0$ such that for this β , the solution of the QP problem of (3) is the same as that of the LPn. Let x be this “common” solution and suppose it has k non-zero components. According to Fuchs [2, 3], this solution must satisfy the following condition (can also be viewed as part of the KKT condition [7])

$$\bar{A}^T(b - \bar{A}\bar{x}) = \beta \text{sgn}(\bar{x}) \quad (4)$$

where \bar{x} is the “reduced solution vector,” made up by the non-zero components of x , \bar{A} is a $m \times k$ matrix made up by the columns of matrix A corresponding to \bar{x} , and $\text{sgn}(\cdot)$ is the usual sign function that for a scalar t

$$\text{sgn}(t) = \begin{cases} -1, & \text{if } t < 0 \\ +1, & \text{if } t > 0 \end{cases} \quad (5)$$

while for a vector v , $\text{sgn}(v)$ is applied component-by-component, leading to a vector of +1s and -1s.

Now, taking the l_2 norm square on both sides of (4), we have

$$(b - \bar{A}\bar{x})^T(\bar{A}\bar{A}^T)(b - \bar{A}\bar{x}) = \beta^2 \|\text{sgn}(\bar{x})\|^2 = \beta^2 k \quad (6)$$

where we used the fact that

$$\|\text{sgn}(\bar{x})\|^2 = (\sqrt{k})^2 = k \quad (7)$$

Note that $\bar{A}\bar{A}^T$ is a $m \times m$ correlation matrix and is semi positive definite. Hence, its eigenvalues are non-negative. Denoting the largest among these as λ_{max} and using the relation between matrix norm and maximum eigenvalues [8], we have

$$(b - \bar{A}\bar{x})^T(\bar{A}\bar{A}^T)(b - \bar{A}\bar{x}) \leq \lambda_{max} \|b - \bar{A}\bar{x}\|^2 \quad (8)$$

Because of eqn (6), we can also write this as

$$\beta^2 k \leq \lambda_{max} \|b - \bar{A}\bar{x}\|^2 \quad (9)$$

As described previously, \bar{x} (and x) is also a solution of the LPn problem, it satisfies the inequality constraint of eqn (2). In fact, based on the results in the Appendix, this solution achieves equality

$$\|b - \bar{A}\bar{x}\|^2 = \delta^2 \quad (10)$$

Hence, we have

$$\beta^2 k \leq \lambda_{max} \delta^2 \quad (11)$$

That is,

$$\beta \leq \sqrt{\frac{\lambda_{max}}{k}} \delta \quad (12)$$

Given δ , this provides an upper bound on β . In practice, we could also use this upper bound as an approximation to β with

$$\beta \simeq \sqrt{\frac{\lambda_{max}}{k}} \delta \quad (13)$$

As demonstrated in Section 3, this approximation can often be quite good. Finally, it is interesting to note that if we let $\delta \rightarrow 0^+$, the LPn problem becomes the LP problem. In this case, the upper bound suggests that $\beta \rightarrow 0^+$, reproducing Fuch's noiseless result for the relation between β and δ (see Section 1).

Using the techniques for deriving the above upper bound, we can also derive a lower bound. However, since $\bar{A}\bar{A}^T$ is semi positive definite rather than positive definite, this is slightly less straightforward and requires some approximations. Specifically, since \bar{A} is an $m \times k$ matrix and since generally, $m > k$ (in practice, m is usually on the order of $5k$ [1]), the rank of matrix $\bar{A}\bar{A}^T$ is at most k . Since $\bar{A}\bar{A}^T$ is an $m \times m$ matrix, it has at most k non-zero eigenvalues. Assume this to be the case and denote the smallest non-zero eigenvalue be denoted as λ_{min} . Let the eigenvectors of $\bar{A}\bar{A}^T$ be e_1, e_2, \dots, e_m , where they are ordered according to the value of their eigenvalues, e_1 for λ_{max} , e_k for λ_{min} , and e_{k+1}, \dots, e_m for 0 eigenvalue. Since $b - \bar{A}\bar{x}$ is an m -dimensional vector, it can be represented by e_1, e_2, \dots, e_m , with

$$b - \bar{A}\bar{x} = \sum_{i=1}^m \alpha_i e_i \quad (14)$$

where α_i are representation coefficients. From this, we have

$$(b - \bar{A}\bar{x})^T (\bar{A}\bar{A}^T) (b - \bar{A}\bar{x}) = \left(\sum_{i=1}^m \alpha_i e_i \right)^T (\bar{A}\bar{A}^T) \left(\sum_{i=1}^m \alpha_i e_i \right) = \sum_{i=1}^m \lambda_i \alpha_i^2 = \sum_{i=1}^k \lambda_i \alpha_i^2 \geq \lambda_{min} \sum_{i=1}^k \alpha_i^2. \quad (15)$$

Now, we find an estimate of $\sum_{i=1}^k \alpha_i^2$. First we notice that the larger sum

$$\sum_{i=1}^m \alpha_i^2 = \|b - \bar{A}\bar{x}\|^2. \quad (16)$$

Hence, from eqn (10) we have

$$\sum_{i=1}^m \alpha_i^2 = \delta^2 \quad (17)$$

Furthermore, we notice that $b - \bar{A}\bar{x} = b - Ax = w$ is the noise. Assume the noise is white (i.e., uncorrelated), on average α_i^2 are roughly the same [12], at δ^2/m . Hence, we have

$$\sum_{i=1}^k \alpha_i^2 \simeq \frac{k}{m} \delta^2 \quad (18)$$

Now, combine eqns (6), (15), and (18), we have

$$\lambda_{min} \frac{k}{m} \delta^2 \leq \beta^2 k \quad (19)$$

That is,

$$\sqrt{\frac{\lambda_{min}}{m}} \delta \leq \beta \quad (20)$$

This provides a lower bound on β for a given δ .

The only problem left now is to find the eigenvalues λ_{min} and λ_{max} . In general these eigenvalues are dependent on the specifics of the A matrix, such as which columns correspond to the non-zero elements of x . However, there is an important case of practical importance where these eigenvalues can be found relatively easily. This is the widely used case where A is an i.i.d. Gaussian random matrix. In this case, it has been shown in previous work [9, 1] that the smallest and the largest eigenvalues for $\bar{A}\bar{A}^T$ are asymptotically

$$\lambda_{min} \simeq m\sigma^2(1 - \sqrt{\gamma})^2, \quad \lambda_{max} \simeq m\sigma^2(1 + \sqrt{\gamma})^2, \quad (21)$$

where m is the number of rows in A and \bar{A} , σ^2 is the variance of each component of A (and \bar{A}), $\gamma = k/m$ (recall that k is the dimension of \bar{x} , also the number of columns of \bar{A}). Plugging these into the bounds of eqns (12) and (20), we then have

$$(1 - \sqrt{\gamma})\sigma\delta \leq \beta \leq \frac{(1 + \sqrt{\gamma})\sigma}{\sqrt{k/m}}\delta \quad (22)$$

When a compressed sensing application uses a Gaussian random sampling matrix, the inequality of (22) can be used to find the range of, or estimate, β from a given δ^2 .

2.2. An Equality Relation

If we add an additional assumption to the derivations in Section 2.1, we can obtain an equality relation between δ and β . Specifically, if we assume that in eqn (4) the $k \times k$ matrix $\bar{A}^T\bar{A}$ is invertible, as Fuchs did in his papers [2, 3], then eqn (4) becomes

$$\bar{x} = (\bar{A}^T\bar{A})^{-1}(\bar{A}^Tb - \beta\text{sgn}(\bar{x})) \quad (23)$$

and this provides a solution to the QP problem. As we mentioned previously, when β matches δ , this solution is the same as that of the LPn. Furthermore, it can be shown that the solution of the LPn must satisfy the inequality with equality (see Appendix). Hence, the solution of (23) should satisfy

$$\|\bar{A}\bar{x} - b\|_2^2 = \delta^2 \quad (24)$$

or

$$\bar{x}^T\bar{A}^T\bar{A}\bar{x} - 2b^T\bar{A}\bar{x} + \|b\|^2 = \delta^2 \quad (25)$$

where we have dropped the subscript 2.

Plugging the right hand side of (23) into (25) and denoting $\text{sgn}(\bar{x})$ as vector c to simplify notation, the first term of (25) becomes

$$\begin{aligned} \bar{x}^T\bar{A}^T\bar{A}\bar{x} &= [(\bar{A}^T\bar{A})^{-1}(\bar{A}^Tb - \beta c)]^T\bar{A}^T\bar{A}[(\bar{A}^T\bar{A})^{-1}(\bar{A}^Tb - \beta c)] \\ &= (\bar{A}^Tb - \beta c)^T[(\bar{A}^T\bar{A})^{-1}]^T(\bar{A}^T\bar{A})[(\bar{A}^T\bar{A})^{-1}(\bar{A}^Tb - \beta c)] \\ &= [(\bar{A}^Tb)^T - \beta c^T][(\bar{A}^T\bar{A})^{-1}(\bar{A}^Tb - \beta c)] \\ &= b^T\bar{A}(\bar{A}^T\bar{A})^{-1}\bar{A}^Tb + c^T(\bar{A}^T\bar{A})^{-1}c\beta^2 - 2b^T\bar{A}(\bar{A}^T\bar{A})^{-1}c\beta \end{aligned} \quad (26)$$

where we have used the fact that $\bar{A}^T\bar{A}$ is symmetrical and so is its inverse, i.e., $[(\bar{A}^T\bar{A})^{-1}]^T = (\bar{A}^T\bar{A})^{-1}$.

Similarly, for the second term of (25), we have

$$\begin{aligned} -2b^T \bar{A}\bar{x} &= -2b^T \bar{A}[(\bar{A}^T \bar{A})^{-1}(\bar{A}^T b - \beta c)] \\ &= -2b^T \bar{A}(\bar{A}^T \bar{A})^{-1} \bar{A}^T b + 2b^T \bar{A}(\bar{A}^T \bar{A})^{-1} c\beta \end{aligned} \quad (27)$$

Combine this and the results of (26) into (25), we have

$$\|\bar{A}\bar{x} - b\|^2 = -b^T \bar{A}(\bar{A}^T \bar{A})^{-1} \bar{A}^T b + c^T (\bar{A}^T \bar{A})^{-1} c\beta^2 + \|b\|^2 = \delta^2 \quad (28)$$

where the terms linear in β (26) and (27) canceled each other. From this, we can find β in terms of δ as:

$$\beta = \sqrt{\frac{\delta^2 - \|b\|^2 + b^T \bar{A}(\bar{A}^T \bar{A})^{-1} \bar{A}^T b}{c^T (\bar{A}^T \bar{A})^{-1} c}} \quad (29)$$

Although this equality provides a more explicit relation between β and δ^2 , in practice it is more difficult to use than the inequalities of Section 2.1 since c and \bar{A} are generally not known before QP and LP are performed.

3. Experimental Verification

In this section, we provide some experimental (simulation) results that verify and illustrate the inequality relations between δ and β derived in Section 2.1. In each experiment, we picked a LPn problem with a given noise level δ^2 [see eqn (2)] and formed a corresponding QP problem [see eqn (3)]. Then, the QP problem was solved for a range of β s and from the resulting solutions, we would try to identify the best β corresponding to the δ since, according to the theory of duality, the best β should result in the same solution as that of the LPn with δ^2 . From this, we could see if, or how well, the best β satisfies the upper and lower bounds derived in Section 2.1. Next, we describe the specific steps in our experiments.

3.1. Experiment Steps

Each experiment consists of the following steps:

1. Generate a sparse random n -dimensional signal x^* .
2. Generate a noisy observed signal $b = Ax^* + w$, where A is an $m \times n$ Gaussian random sampling matrix with $m < n$ and component variance σ^2 and w is an m -dimensional additive white Gaussian noise vector with variance σ_w^2 .
3. Reconstruct x^* by solving the LPn problem with δ^2 set to $\delta^2 = m\sigma_w^2$. Denote the resulting solution as $\hat{x}(\delta)$.
4. Reconstruct x^* by solving the QP problem with a range of β s. Denote the resulting solution as $\hat{x}(\beta)$.
5. Compare the minimum obtained by LPn, i.e., $\|\hat{x}(\delta)\|$, and the maximum of the dual function obtained by QP, $g(\beta)$ (more details later).

6. Compare the normalized reconstruction errors (i.e., $\|x^* - \hat{x}\|_2/\|x^*\|$) obtained by LPn and QP (\hat{x} could be either $\hat{x}(\delta)$ or $\hat{x}(\beta)$).
7. Find “the best β ” (that produces the same QP solution as the LPn solution) and compare this β with the bounds in eqns (12) and (20).

We now explain each of these steps in some detail. In Step 1, the original random sparse signal x^* was obtained from examples provided/generated by the L1 Magic software [10] (which uses these to illustrate the workings of compressed sensing algorithms). Specifically, in our experiments x^* is a sparse random vector of dimension $n = 256$ and its non-zero components consists of $k = 24$ randomly placed +1s and -1s (also randomly chosen), as shown in Fig. 2.

In Step 2, the noisy observed signal b was generated using a $m \times n$ Gaussian random sampling matrix A with $m = 100$ and component variance $\sigma^2 = 1$; for the additive noise w , we used a white Gaussian noise with variance σ_w^2 (whose value is different in different experiments, more details later). Some typical noisy observed signals are also shown in Fig. 2.

In Step 3, the LPn problem was solved with a log barrier algorithm in L1 Magic and in Step 4, the QP problem was solved with the L1 Regularization software developed by Kim et al [11].

In Step 5, what we are really doing is to use results of the duality theory [7] to find the best β . Specifically, for the LPn problem, one can define a dual function

$$g(\lambda) = \inf_u \left\{ \|u\|_1 + \lambda(\|Au - b\|_2^2 - \delta^2) \right\} \quad (30)$$

Because the LPn satisfies the strong duality condition (see [7]), we have

$$\|\hat{x}\|_1 \geq g(\lambda) \quad \text{for all } \lambda \geq 0 \quad (31)$$

where \hat{x} is the solution of the LPn problem and equality is achieved at the best λ , denoted as λ_0 , with

$$\lambda_0 = \arg \max g(\lambda) \quad \text{and} \quad \|\hat{x}\|_1 = g(\lambda_0) \quad (32)$$

Now, the dual function $g(\lambda)$ can be linked to the QP solution: we can re-write it as

$$g(\lambda) = \inf_u \left\{ \|u\|_1 + \lambda \left(\|Au - b\|_2^2 - \delta^2 \right) \right\} = 2\lambda \inf_u \left\{ \frac{1}{2} \|Au - b\|_2^2 + \frac{1}{2\lambda} \|u\|_1 \right\} - \lambda\delta^2 \quad (33)$$

where the $\inf_x \{\cdot\}$ is the QP solution with $1/2\lambda = \beta$. In this way, whenever a QP problem with β is solved, we can compute a corresponding $g(\lambda)$ with $\lambda = 1/2\beta$. In this sense, we can write (re-parameterize) $g(\lambda)$ as $g(\beta)$ and the strong duality of eqn (32) can be re-written in terms of β as

$$\beta_0 = \arg \max g(\beta) \quad \text{and} \quad \|\hat{x}\|_1 = g(\beta_0) \quad (34)$$

where β_0 is the best β (which makes the QP having the same solution as that of the LP).

Finally, Steps 6 and 7 are relatively straightforward and we will discuss our experimental results next.

3.2. Experimental Results

Some typical experimental results are shown in Figs. 3-14. In Fig. 3, the observation noise variance is $\sigma_w^2 = 0.0225$, corresponding to an SNR of 30dB (low noise) and a noise bound of $\delta^2 = m\sigma_w^2 = 100 \times 0.0225 = 2.25$. Fig. 3 contains information obtained in Step 5, i.e., the minimum achieved by the LPn ($\|\hat{x}\|_1$), the re-parametized dual function $g(\beta)$, and the upper and lower bounds for β we derived in Section 2.1. Note that since the minimum achieved by LPn is a number (constant) while the dual function is a function of β , we presented the former as a constant line. Similarly, the bounds are numbers, i.e., specific values of β , hence are presented as vertical lines.

From the results of Fig. 3, we can make two observations. First, the experimental results agrees with the prediction of the duality theory. That is, the $g(\beta)$ curve is always below the $\|\hat{x}\|_1$ line and for the “best” β , the curve approaches the line. Second, the best β falls in an interval predicted/defined by our upper and lower bounds (almost right in the middle of the interval). This is very encouraging.

Fig. 4 compares the normalized reconstruction errors (see Step 6) for the LPn and QP. The former is a number, presented as a horizontal line while the latter is a function of β , hence is a curve. At the “best” β [i.e., when eqn (34) is satisfied or, when the curve meets the straight line in Fig. 3], the QP and LP have the same reconstruction errors. Hence, looking at reconstruction errors of QP and LP provides another potential way¹ to identifying the “best” β , as can be seen in Fig. 4; for the “best” β , the QP has the same reconstruction error as that of LP. From Fig. 4, we can also observe that β can have a strong effect on reconstruction error. Finally, we note that the best β does not necessarily lead to minimum reconstruction error for QP since the best β is best in the sense of duality theory (providing the same solution as that of LPn), not in the sense of minimum reconstruction error. Currently, it is not obvious as to how to find the best β in this latter sense.

To ensure that our results in Figs. 3 and 4 are no accidents, we repeated that experiment 100 times (each time with a new random sparse signal, random sampling matrix, and additive noise vector) and averaged their results. These are shown in Figs 5 and 6. As can be seen, the nature of the results are the same as that of Fig. 3 and 4.

Finally, we repeated the experiment for Figs. 3-6 for a higher noise level, with $\sigma_w^2 = 0.2025$, corresponding to an SNR of 20dB (heavy noise) and a noise level of $\delta^2 = 100 \times 0.2025 = 20.25$. The results are presented in Figs. 7-10. The nature of the results is the same as that of Figs. 3-6: our derived bounds worked well. Furthermore, compared with Figs 3-6, we can see that as the noise reduces (from 20dB to 30dB SNR), the bounds and the best β move to the left, agreeing with the theoretical prediction that the best $\beta \rightarrow 0^+$ when the noise level reduces to 0.

Appendix

Consider the solution to the ℓ_1 minimization problem

$$(P_1) \quad \min \|u\|_1 \quad \text{subject to} \quad \|Au - b\|_2 \leq \delta.$$

First we note that if $\|b\|_2 \leq \delta$, then $\hat{x} = 0$ is the solution to (P_1) . To avoid this trivial case, we assume $\|b\|_2 > \delta$.

The following result belongs to a well-known result in convex optimization, known as the maximum principle. We include it here for the reader’s convenience. Plus, our proof is specialized to the (P_1) problem.

¹Sometimes, the normalized QP error curve can intersect the LP error line at more than one β , in this case, we need to rely on the dual function to identify the best β .

Maximum Principle: Let \hat{x} be a minimizer of (P_1) and if $\hat{x} \neq 0$, then

$$\|A\hat{x} - b\|_2 = \delta.$$

Proof: In fact, since $\hat{x} \neq 0$, we may assume $\hat{x}(i_0) \neq 0$ for some i_0 ($x(i)$ is the i th component of x). Suppose \hat{x} is not on the boundary, then

$$d = \delta - \|A\hat{x} - b\|_2 > 0.$$

Choose a small t such that $|\hat{x}(i_0) - t| < |\hat{x}(i_0)|$ and $\|A(\hat{x} - \hat{x}')\|_2 < d$ (here $\hat{x}'(i) = \hat{x}(i)$ for $i \neq i_0$ and $\hat{x}'(i_0) = \hat{x}(i_0) - t$), we get a contradiction because $\|A\hat{x}' - b\|_2 < \delta$ and $\|\hat{x}'\|_1 < \|\hat{x}\|_1$. This proves the maximum principle.

References

- [1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. IT*, Vol. 51, pp. 4203-4215, Dec. 2005.
- [2] J. J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. IT*, Vol. 50, pp. 1341-1344, June 2004.
- [3] J. J. Fuchs, "Recovery of exact sparse representations in the presence of bounded noise," *IEEE Trans. IT*, Vol. 51, pp. 3601-3608, Oct. 2005.
- [4] J. A. Tropp, "Just relax: convex programming methods for identifying sparse signals," *IEEE Trans. Info. Theory*, Vol. 51, pp. 1030-1051, Mar. 2006.
- [5] J. Tang, B. Nett, and G.-H. Chen, "Performance comparison between compressed sensing and statistical iterative reconstruction," *Proc. SPIE Medical Imaging Conf*, 2009.
- [6] <http://www-stat.stanford.edu/tibs/lasso.html>
- [7] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [9] S. Geman, "A limit theorem for the norm of random matrices," *Annals of Probability*, Vol. 8, pp. 253-261, 1980.
- [10] L1-Magic, <http://www.acm.caltech.edu/l1magic/>.
- [11] L1-Regularization, http://www.stanford.edu/~boyd/l1_ls/.
- [12] H. L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley and Sons, 1968.

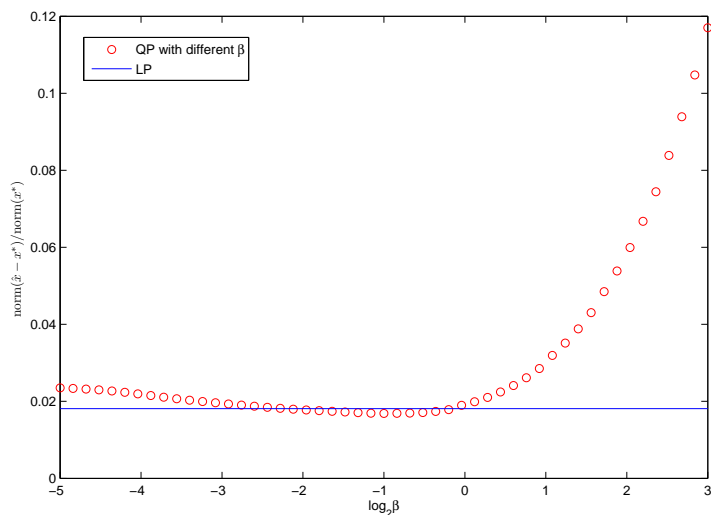


Figure 1: QP's Reconstruction Results as a Function of β . Curve: QP's reconstruction error as a function of β ; Straight line: LP's reconstruction error. $n = 100, k = 10, m = 50, \delta^2 = 0.75$ (SNR = 37dB)

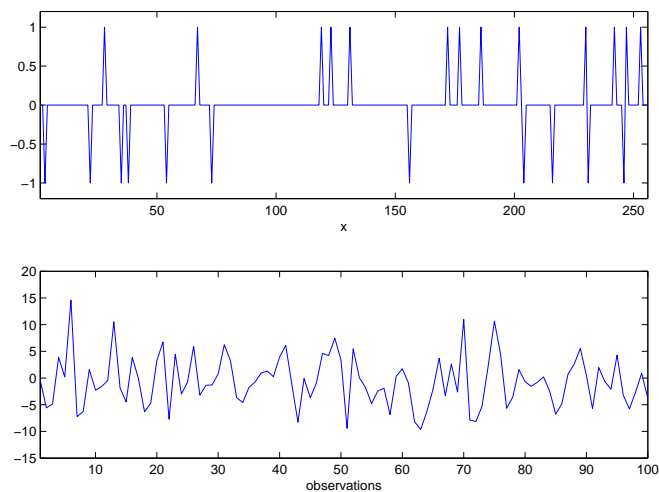


Figure 2: A Sparse Random Signal and its Noisy Observation. Top: the sparse random signal, bottom: the noisy observed signal with an SNR of 20dB.

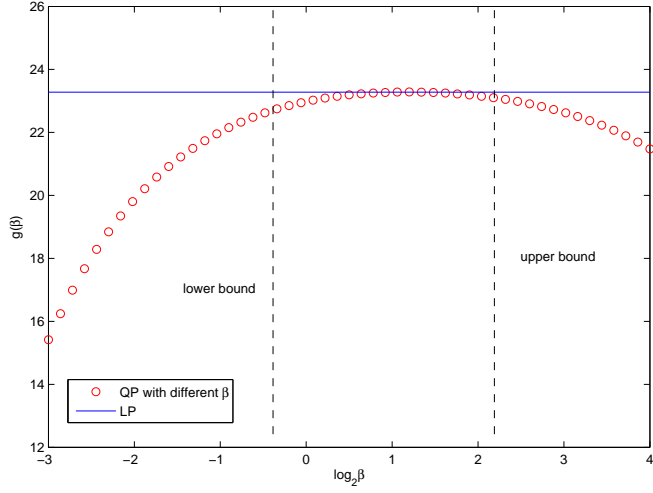


Figure 3: The Minimum from LP and the Dual Function $g(\beta)$ from QP: the 30dB SNR Case. Straight line: the minimum from LP, curve: the dual function from QP.

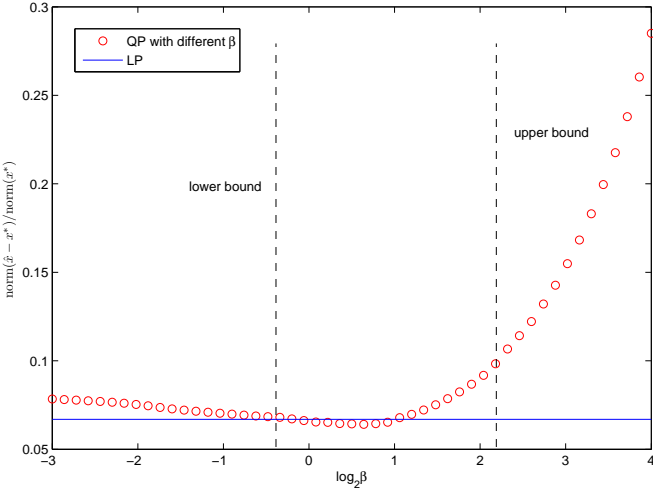


Figure 4: Recovery Error from LP and QP: the 30dB SNR Case. The straight line: LP, the curve: QP.

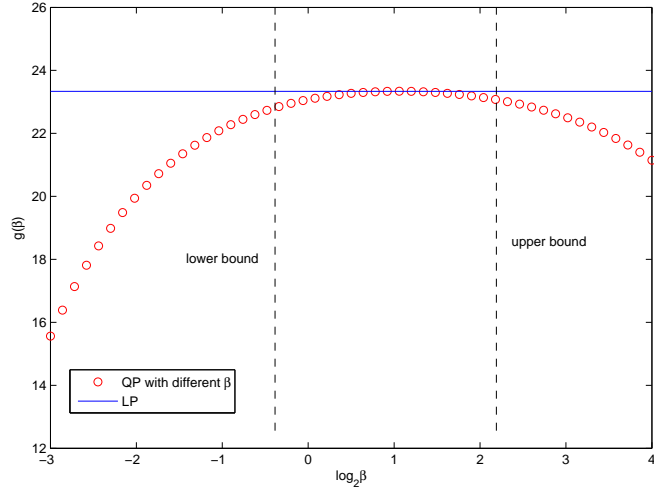


Figure 5: Experiment in Figs. 3 and 4 Repeated 100 Times: Averaged Minimum from LP and Averaged Dual Function from QP.

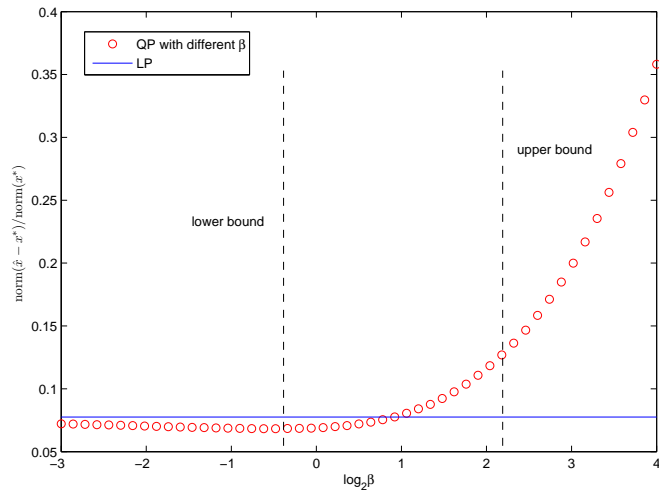


Figure 6: Experiment in Figs. 3 and 4 Repeated 100 Times: Averaged Reconstruction Errors from LP and QP.

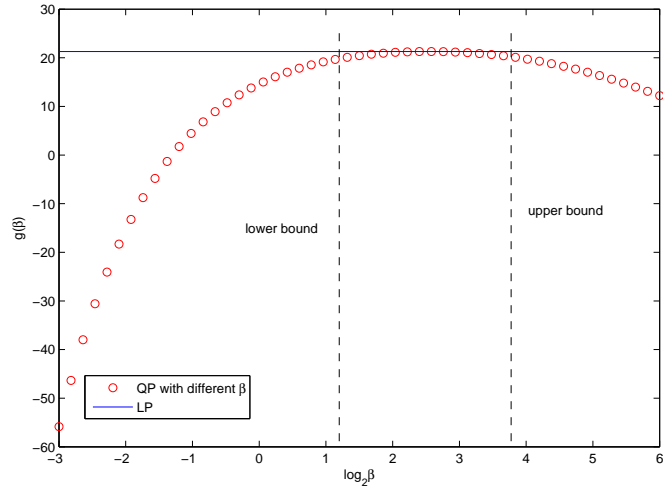


Figure 7: The Minimum from LP and the Dual Function $g(\beta)$ from QP: the 20dB SNR Case

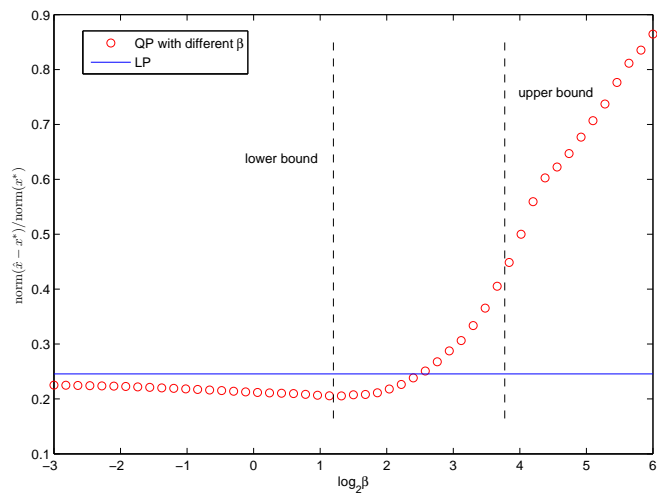


Figure 8: Recovery Error from LP and QP: the 20dB SNR Case. The straight line: LP, the curve: QP.

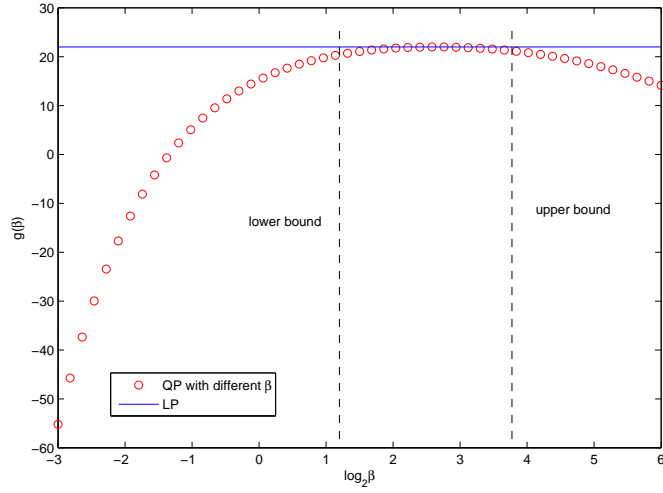


Figure 9: Experiment in Figs. 7 and 8 Repeated 100 Times: Averaged Minimum from LP and Averaged Dual Function from QP.

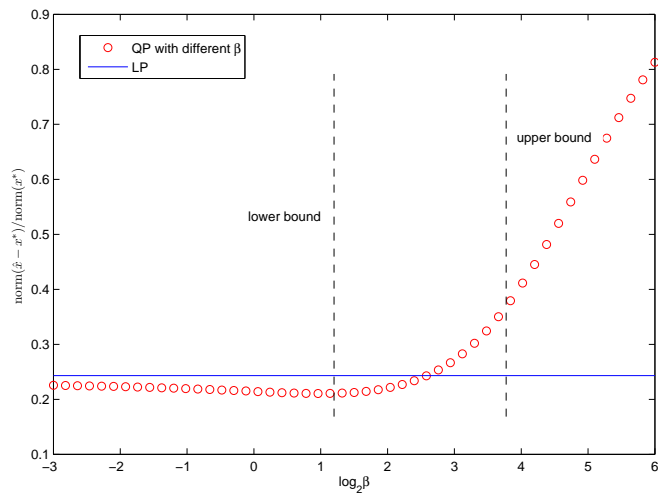


Figure 10: Experiment in Figs. 7 and 8 Repeated 100 Times: Averaged Reconstruction Errors from LP and QP.